

Université de Sherbrooke

Le contrôle de l'épissage alternatif par les protéines hnRNP H et hnRNP A1

Par

Jean-François Fisette

Département de microbiologie et d'infectiologie

Thèse présentée à la Faculté de médecine et des sciences de la santé

En vue de l'obtention du grade de

philosophiae doctor (Ph.D.) en microbiologie

22 avril 2009

Dr Benoit Chabot, microbiologie (évaluateur interne)

Dr Antonio Conconi, microbiologie (évaluateur interne)

Dr Éric Massé, biochimie (évaluateur interne)

Dr François Dragon, département des sciences biologiques,

Université du Québec à Montréal (évaluateur externe)



Library and Archives
Canada

Published Heritage
Branch

395 Wellington Street
Ottawa ON K1A 0N4
Canada

Bibliothèque et
Archives Canada

Direction du
Patrimoine de l'édition

395, rue Wellington
Ottawa ON K1A 0N4
Canada

Your file Votre référence
ISBN: 978-0-494-52831-0
Our file Notre référence
ISBN: 978-0-494-52831-0

NOTICE:

The author has granted a non-exclusive license allowing Library and Archives Canada to reproduce, publish, archive, preserve, conserve, communicate to the public by telecommunication or on the Internet, loan, distribute and sell theses worldwide, for commercial or non-commercial purposes, in microform, paper, electronic and/or any other formats.

The author retains copyright ownership and moral rights in this thesis. Neither the thesis nor substantial extracts from it may be printed or otherwise reproduced without the author's permission.

AVIS:

L'auteur a accordé une licence non exclusive permettant à la Bibliothèque et Archives Canada de reproduire, publier, archiver, sauvegarder, conserver, transmettre au public par télécommunication ou par l'Internet, prêter, distribuer et vendre des thèses partout dans le monde, à des fins commerciales ou autres, sur support microforme, papier, électronique et/ou autres formats.

L'auteur conserve la propriété du droit d'auteur et des droits moraux qui protègent cette thèse. Ni la thèse ni des extraits substantiels de celle-ci ne doivent être imprimés ou autrement reproduits sans son autorisation.

In compliance with the Canadian Privacy Act some supporting forms may have been removed from this thesis.

While these forms may be included in the document page count, their removal does not represent any loss of content from the thesis.

Conformément à la loi canadienne sur la protection de la vie privée, quelques formulaires secondaires ont été enlevés de cette thèse.

Bien que ces formulaires aient inclus dans la pagination, il n'y aura aucun contenu manquant.


Canada

TABLE DES MATIÈRES

Table des matières.....	I
Liste des figures.....	III
Liste des abréviations.....	IV
Résumé.....	VI
Introduction.....	1
L'épissage alternatif et son importance.....	1
L'épissage constitutif.....	3
Le contrôle de l'épissage alternatif.....	9
Les protéines hnRNP H et hnRNP F.....	21
Épissage alternatif de l'ARN pré-messager de hnRNP A1.....	23
Chapitre 1.....	28
*Martinez-Contreras R., *Fisette J-F., *Nasim F., Cordeau M., Chabot B., Intronic binding sites for hnRNP A/B and hnRNP F/H proteins stimulate pre-mRNA splicing. <i>PLoS Biol</i> , 2006. 4(2): p. e21. (*ces auteurs ont contribué également).	
Résumé.....	29

Chapitre 2.....	78
 Fisette J-F., Toutant J., Dugré-Brissou S., DesGroseillers L. and Chabot B. hnRNP A1 and hnRNP H can collaborate to modulate 5' splice site selection, (article soumis dans <i>RNA</i>).	
Résumé.....	79
Discussion.....	123
L'activité des sites de liaison pour hnRNP H	123
Interactions homotypiques et hétérotypiques.....	131
Génomique humaine et modulation de l'épissage.....	134
Perspectives d'implications des autres régulateurs d'épissage.....	136
Application de notre modèle.....	138
Conclusions.....	141
Remerciements.....	142
Références.....	143

LISTE DES FIGURES

Figure 1.....	4
Figure 2.....	5
Figure 3.....	8
Figure 4.....	10
Figure 5.....	14
Figure 6.....	18
Figure 7.....	21
Figure 8.....	25
Figure 9.....	125
Figure 10.....	131
Figure 11.....	135
Figure 12.....	139

LISTE DES ABRÉVIATIONS

VEGF : «vascular endothelial growth factor »

NRXN : «neurexin»

snRNP: « small nuclear ribonucleoproteins »

U2AF: «U2 snRNP auxiliary factor»

WT1: «Wilms tumor»

hnRNP: « heterogeneous nuclear ribonucleoprotein»

hnRNA: « heterogeneous nuclear RNA »

DSCAM: « down syndrome cell adhesion molecule »

RRM : « RNA recognition motif »

GRD : « glycine-rich domain »

CstF-64 : « cleavage stimulation factor-64 »

STAR: « signal transduction and activation of RNA »

NMD : « nonsense-mediated decay »

FMR : « fragile X mental retardation »

FMRP : « fragile X mental retardation protein »

BRET : « bioluminescence resonance energy transfert »

YFP : « yellow fluorescence protein »

SMN : « survival motor neuron »

RÉSUMÉ

Les protéines hnRNP A1 sont impliquées dans l'épissage alternatif. Un mode d'action proposé implique la formation d'homodimères entre molécules hnRNP A1 causant un réarrangement dans la structure de l'ARN pré-messager. Cette modulation de l'ARN permettrait le rapprochement de sites d'épissage 5' et 3' d'exons situés de par et d'autres d'un exon alternatif. Le domaine riche en résidus glycines est responsable, en grande partie, de l'interaction entre les deux protéines hnRNP A1. Comme la protéine hnRNP H contient aussi un domaine riche en résidus glycines, nous avons postulé que cette dernière pouvait moduler l'épissage alternatif de la même manière que hnRNP A1. Afin de vérifier cette hypothèse, nous avons utilisé un ARN pré-messager constitué de deux sites d'épissage 5' (distal et proximal) en compétition pour un seul site d'épissage 3'. En présence de sites de liaison pour hnRNP H, nous observons que le choix du site d'épissage 5' est déplacé vers le site distal. Nous avons confirmé le rôle des protéines hnRNP H dans la sélection des sites d'épissage 5' *in vitro* et avons déterminé que le domaine riche en résidus glycines (GRD) est important pour l'activité d'épissage de ce régulateur. Nous avons ensuite exploré la possibilité que des combinaisons de sites de liaison pour hnRNP H et hnRNP A1 puissent activer l'utilisation du site d'épissage 5' distal. Nous avons observé que des combinaisons hétérotypiques peuvent reproduire cette activité d'épissage. Finalement, nous avons utilisé la technologie BRET (« bioluminescence resonance energy transfer ») pour démontrer que des interactions homotypiques entre protéines hnRNP H et hétérotypiques entre molécules hnRNP A1 et hnRNP H peuvent se former dans les cellules vivantes. Notre étude suggère que les

protéines hnRNP H et hnRNP A1 peuvent changer la conformation de l'ARN pré-messager et affecter le choix du site d'épissage.

Mots clés : épissage alternatif, site d'épissage 5' en compétition, hnRNP A1 et hnRNP H

INTRODUCTION

L'épissage alternatif et son importance

La transcription des gènes eucaryotes par l'ARN polymérase II est suivie et même accompagnée d'une série d'événements de maturation des ARN pré-messagers qui génèrent des transcrits matures qui sont transportés au cytoplasme pour ensuite être traduits. Parmi ces modifications, l'épissage est le processus qui permet d'exciser certaines régions (introns) et de joindre les exons. Une variante de ce mécanisme, appelée épissage alternatif, modifie l'inclusion ou l'exclusion de certaines portions d'un ARN pré-messenger et engendre une variété de transcrits codant pour différentes protéines. Les dernières études ont estimé que chez l'humain, 92 à 95% des gènes seraient épissés de façon alternative (Pan *et al.*, 2008; Wang *et al.*, 2008). Étant donné que la quasi-totalité des 25 000 gènes humains (Pennisi, 2005) est soumise à l'épissage alternatif, ce processus cellulaire serait donc l'élément clé permettant d'augmenter le réservoir protéique.

L'épissage alternatif permet d'intégrer ou non certaines séquences codantes pour des domaines essentiels à l'activité d'une protéine. Dans certaines situations, des isoformes produits à partir d'un même gène peuvent avoir des activités antagonistes et gérés des processus biologiques très importants. Par exemple, dans le cas de la mort cellulaire programmée (apoptose), l'épissage alternatif du gène *Bcl-x* peut générer un isoforme soit pro- (*Bcl-x_S*) ou anti-apoptotique (*Bcl-x_L*) (Boise *et al.*, 1993). Le facteur de croissance VEGF (« vascular endothelial growth factor »), impliqué dans l'angiogénèse, est aussi épissé de façon alternative: une molécule pro- (VEGF-A) ou anti-angiogénique (VEGF-b) peut être produite afin de contrôler la formation de vaisseaux sanguins

(Ladomery et al., 2007). Le nombre important d'isoformes produit à partir d'un même gène peut avoir des conséquences biologiques profondes sur le développement des organismes. Les 500 isoformes générés à partir de l'épissage alternatif du gène *slo* ont un impact primordial sur le développement de l'oreille interne. En effet, plusieurs des protéines produites sont impliquées dans la formation du limaçon et, de par leurs propriétés physiologiques différentes, celles-ci permettent de capter avec précision une large gamme de fréquences (Graveley, 2001). Le développement du système nerveux est aussi contrôlé par des décisions d'épissage alternatif : à partir des trois gènes d'adhésion cellulaire *NRXN1*, *NRXN2* et *NRXN3* (« *neurexin* »), 2250 isoformes peuvent être produits et les protéines traduites à partir de ceux-ci ont un rôle primordial dans la connectivité spécifique des neurones (Rowen et al., 2002).

Étant donné que l'épissage alternatif peut générer une quantité importante de transcrits, codant pour une multitude de protéines aux fonctions diverses, des dérèglements dans les niveaux relatifs d'isoformes peuvent être catastrophiques pour une cellule. Des mutations ciblant directement des séquences régulatrices des ARN pré-messagers peuvent modifier des décisions d'épissage alternatif et engendrer plusieurs maladies graves. Par exemple, chez les patients atteints du syndrome de Frasier, une série de mutations en aval de l'exon 9 du gène suppresseur de tumeurs « *Wilms* » (*WT1*) serait responsable de la pathologie en inactivant l'utilisation d'un site d'épissage 5' alternatif et entraînant la production d'une protéine dont trois acides aminés essentiels sont exclus. (Faustino et Cooper, 2003). L'expression d'un dominant négatif du gène de l'hormone de croissance (*GH-I*) serait impliquée dans la déficience familiale isolée en hormone de croissance type II. C'est un changement dans la sélection d'un site d'épissage 3' alternatif

(causé par une série de mutations en aval et en amont du site) qui engendre l'exclusion d'une région importante du transcrit : l'ARN messenger qui en résulte codera pour une protéine dominante-négative (Faustino et Cooper, 2003). Plusieurs événements d'épissage alternatif ont été associés à une variété de cancers, suggérant l'implication de ceux-ci dans l'apparition de la maladie. On note par exemple que chez le cancer du poumon, un mauvais patron d'épissage alternatif du gène *actinine-4* résulte en une protéine dont la localisation cellulaire est modifiée et, conséquemment, l'intégrité du cytosquelette des cellules pulmonaires est affectée (Honda et al., 2004). Étant donné que les cellules cancéreuses semblent avoir des profils d'épissage alternatif qui les distinguent des cellules normales, la mesure de ratios de profils d'épissage alternatif de gènes clés permet de générer des marqueurs moléculaires. C'est le cas du cancer du sein où le profil d'épissage alternatif de 600 gènes a été vérifié et permis d'identifier 41 événements spécifiques aux cellules cancéreuses (Venables et al., 2008a).

L'épissage constitutif

L'excision des introns et la juxtaposition des exons impliquent deux réactions de transestérification qui sont effectuées à des séquences consensus nommées sites d'épissage. Chez l'humain, à la première extrémité de l'intron est retrouvé le site d'épissage 5' défini par la séquence MAG/GURAGU (/ indique le site d'épissage, M étant A ou C et R étant A ou G) (Mount *et al.*, 1983). L'autre bout de l'intron est quant à lui composé du site d'épissage 3', caractérisé par trois régions distinctes : le site de branchement YNYURAC (où Y = C ou T et N = G,A,T ou C) (Will *et al.*, 2001), une séquence riche en pyrimidines et le site consensus d'épissage 3' qui est constitué de la séquence YAG/G (/ indique le site d'épissage et Y = C ou T) (Mount *et al.*, 1983). Les

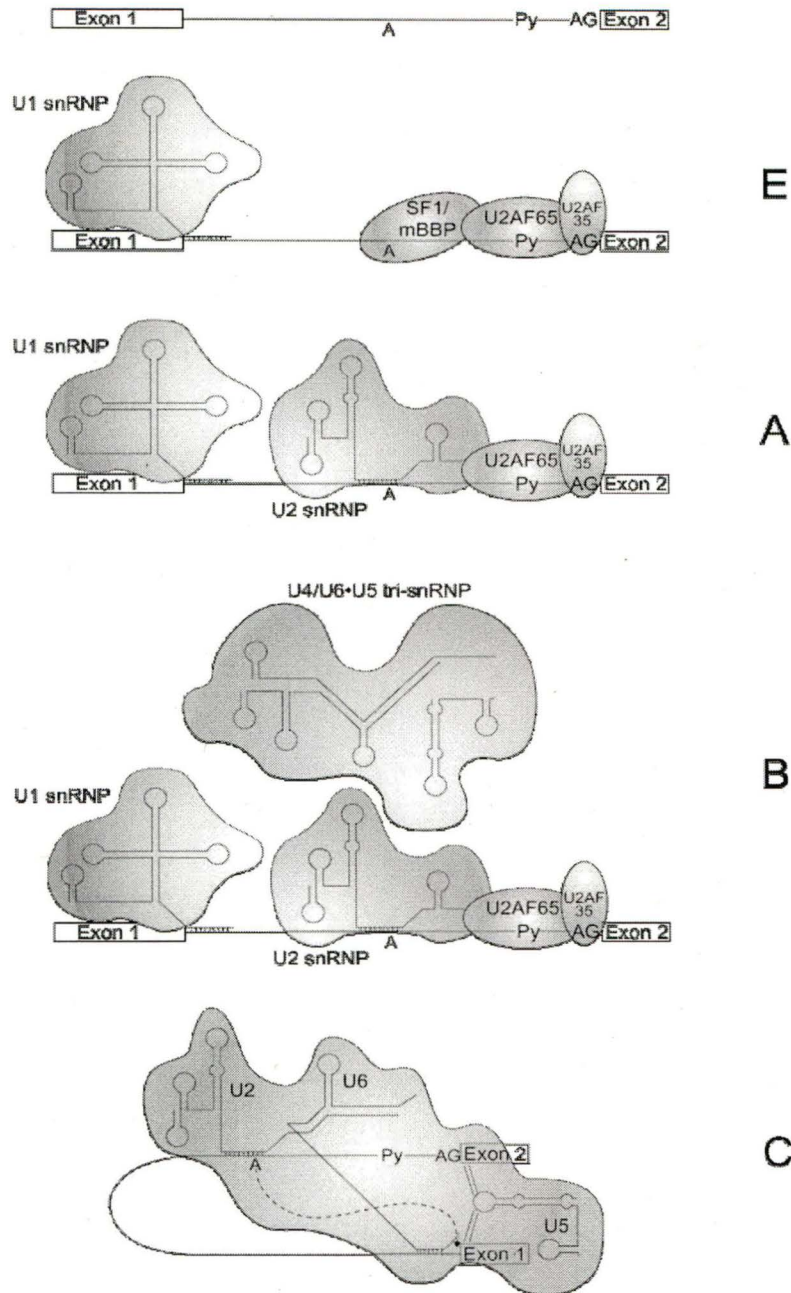


Figure 1. Assemblage du « spliceosome ». La première étape dans l'assemblage consiste en la formation du complexe E correspondant à la liaison de la snRNP U1 au site d'épissage 5', la reconnaissance du site de branchement par la protéine SF1 et l'arrivée des composantes U2AF65 et U2AF35 à la séquence riche en pyrimidines et au dinucléotide AG respectivement. Le passage au complexe A s'effectue lorsque la snRNP U2 se lie au site de branchement. Se joint ensuite la tri-snRNP U4/U5/U6 pour former le complexe B. La snRNP U4 est éjectée et la snRNP U6 remplace la snRNP U1 dans le complexe C : c'est à ce moment qu'aura lieu la première des deux réactions de transestérification. Tiré de (Graveley, 2000).

deux réactions d'épissage sont catalysées par une immense macromolécule nommée « spliceosome » qui est composée de cinq particules « small nuclear ribonucleoproteins » (snRNP) et de protéines accessoires. Plusieurs étapes d'assemblage du « spliceosome » aux sites d'épissage et de changements de conformation de l'ARN pré-messager sont nécessaires afin de catalyser les deux réactions (figure 1). Tout d'abord, la snRNP U1 est responsable de la reconnaissance du site d'épissage 5' par l'intermédiaire de sa composante ARN (Rosbash et Seraphin, 1991). Le site de branchement est reconnu par la protéine SF1, la région riche en pyrimidines est liée par U2AF65 (« U2 snRNP Auxiliary Factor 65 kDa») alors que le dinucléotide AG au site d'épissage 3' est reconnu par U2AF35 (Brow, 2002). Une interaction entre les protéines SF1 et U2AF65 engendre une coopération dans la liaison des facteurs à leurs sites respectifs et contribuerait à une meilleure reconnaissance du site de branchement (Berglund et al., 1998). Cette première

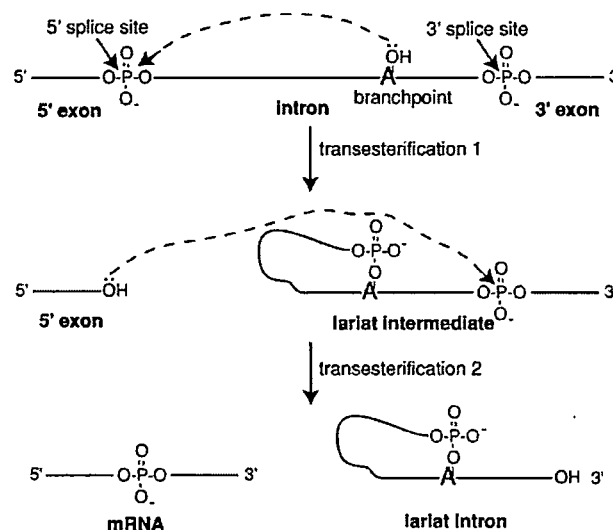


Figure 2. Les deux réactions de transestérification. Dans la première réaction, le groupement OH en 2' de l'adénosine forme un lien avec le phosphate au site d'épissage 5'. Un intermédiaire sous forme de lasso est formé suite à cette réaction. Par la suite, le groupement 3' hydroxyle libre du premier exon attaque le phosphate à l'extrémité 5' du deuxième exon. Cette dernière étape lie les deux exons et relâche la structure en lasso. Tiré de (Brow, 2002).

étape dans l'assemblage du « spliceosome » forme le complexe E. Par la suite, le passage au complexe A est caractérisé par l'arrivée de la snRNP U2 au site de branchement (grâce à sa composante ARN). Lorsque la tri-snRNP U4/U5/U6 se joint aux composantes, il y a formation du complexe B. À cette étape, l'ARN pré-messager subit d'importants changements de conformation et passe au complexe C. Cette étape est caractérisée par le remplacement de la snRNP U1 par la snRNP U6 et l'éjection de la snRNP U4. C'est par la suite que les deux réactions de transestérification sont catalysées par le « spliceosome » (figure 2). La première consiste en la formation d'une structure en lasso par l'attaque du groupement 2' hydroxyle de l'adénosine (situé au site de branchement) au phosphate de l'extrémité 5' de l'intron. La deuxième s'effectue lorsque le groupement hydroxyle libre en 3' du premier exon attaque le phosphate à l'extrémité 5' du deuxième exon : cette réaction permet de liguer les deux exons et de libérer la structure en lasso (Black, 2003). Ces étapes dans l'assemblage du « spliceosome » ont été caractérisées par différentes approches biochimiques permettant d'identifier la composition en snRNP des macromolécules et les différences de mobilité des complexes dans un gel natif ou un gradient de densité (Jurica et Moore, 2003). Cependant, les difficultés techniques liées à la stabilité des complexes rendent ardues l'établissement de modèles précis. Conséquemment, certaines variations dans l'assemblage séquentiel des snRNP ont été proposées (Brow, 2002).

L'épissage des ARN pré-messagers doit s'effectuer de façon très précise afin d'éviter toutes erreurs pouvant résulter en un transcrit codant pour une protéine dont l'activité serait modifiée. Par exemple, l'insertion d'une portion de l'ARN pré-messager pourrait inclure une section qui serait en phase avec le reste du transcrit et qui coderait

pour un domaine supplémentaire. Chez l'humain, le « spliceosome » doit s'assurer de juxtaposer les bons sites d'épissage dans un contexte où les séquences qui les définissent sont très variables (Sun et Chasin, 2000). De plus, cet organisme possède des introns dont la taille est souvent immense. En effet, celle-ci peut varier de 60 à 200 000 nucléotides et 20 % des introns sont plus grands que 5000 nucléotides (Lander et al., 2001). Les exons sont quant à eux beaucoup plus petits avec une taille moyenne de 145 nucléotides (Lander et al., 2001). Or, dans cette mer de nucléotides, le « spliceosome » doit distinguer les sites authentiques des pseudos sites (séquences correspondant parfaitement à ceux des sites d'épissage sans toutefois qu'elles ne soient utilisées) (Sun et Chasin, 2000). Le cas du gène de la *dystrophine* illustre une situation où la machinerie d'épissage accomplit un travail hautement précis: un transcrit supérieur à deux millions de nucléotides est généré, comporte 78 introns à épisser, dont la taille moyenne de ceux-ci est de 26 137 nucléotides (Pozzoli et al., 2002).

Afin que le « spliceosome » engage les bons sites d'épissage dans les deux réactions de transestérification, des mécanismes cellulaires ont été proposés pour expliquer la bonne sélection des séquences et l'inhibition des pseudos sites. Dans un premier temps, certaines protéines ne faisant pas partie de la machinerie d'épissage constitutif peuvent contribuer à l'identification des bons sites. Par exemple, l'interaction de la protéine DEK avec U2AF35 assure la bonne sélection du dinucléotide AG au site d'épissage 3' et prévient l'assemblage de l'unité U2AF65 à une région riche en pyrimidines qui n'est pas suivie de ce dinucléotide (Soares et al., 2006). D'un autre côté, les pseudos sites sont souvent bordés de séquences ayant la propriété d'inactiver leur utilisation. Dans certains cas, ces motifs inhibiteurs correspondent à des séquences de

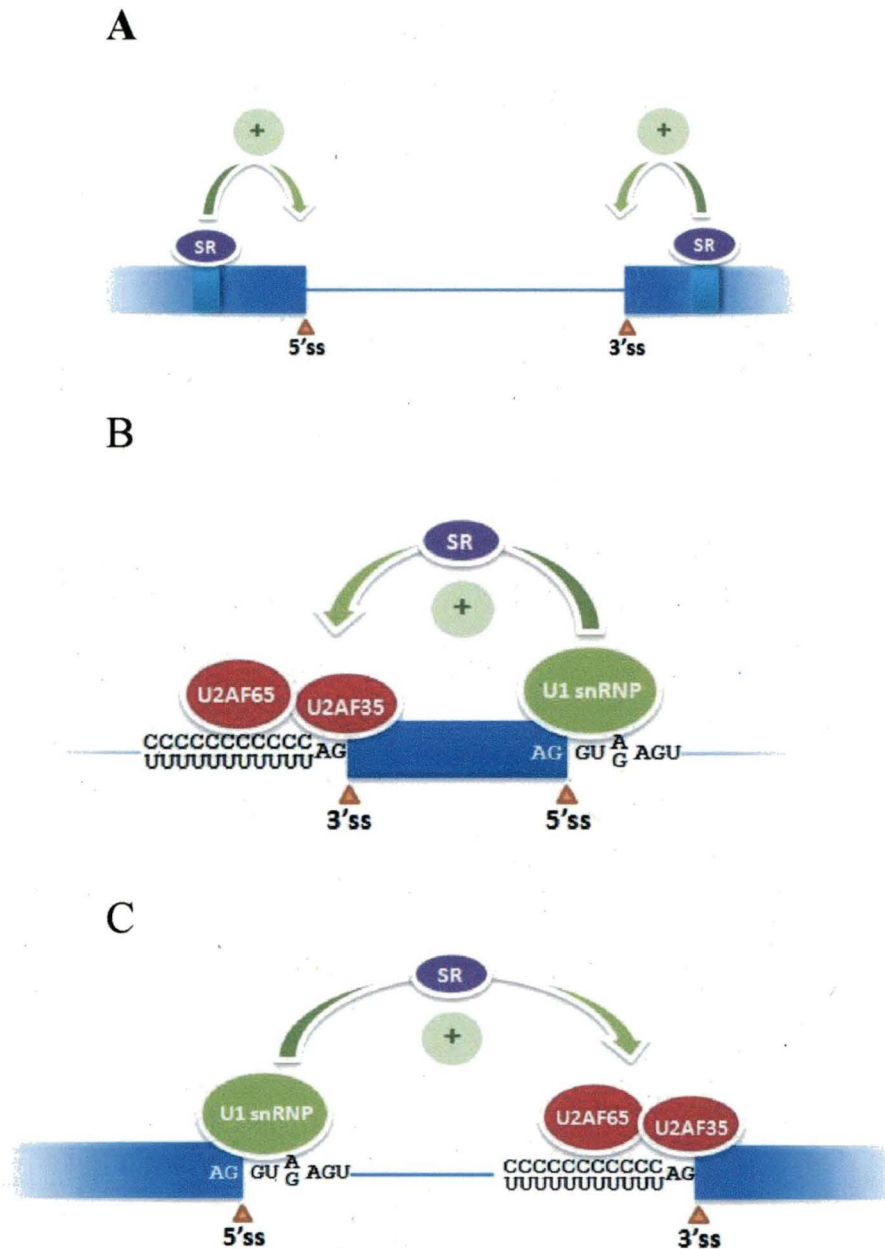


Figure 3. A) Activation des sites d'épissage 5' et 3' par les protéines SR. Des séquences exoniques peuvent recruter des protéines SR et faciliter l'assemblage du « spliceosome ». **B) La définition de l'exon.** La liaison de la snRNP U1 au site d'épissage 5' peut faciliter l'assemblage des complexes au site d'épissage 3' par l'intermédiaire des protéines SR. **C) La définition de l'intron.** Les protéines SR permettent la communication entre la composante protéique de la snRNP U1 et U2AF35 le long de l'intron. Tiré et modifié de (Fisette *et al.*, 2009).

reconnaissance pour certains facteurs protéiques, dont hnRNP H, qui fera l'objet d'une description plus détaillée dans les prochaines pages. Plusieurs études ont d'ailleurs démontré que ces éléments inhibiteurs sont retrouvés plus fréquemment près des pseudos sites qu'à proximité des sites authentiques (Sironi *et al.*, 2004; Wang *et al.*, 2004; Zhang et Chasin, 2004). À l'inverse, des séquences activatrices assurent l'utilisation de certains sites. C'est le cas des sites d'épissage authentiques dont le consensus est faible : ceux-ci contiennent une plus grande fréquence d'éléments positifs à proximité qui contribue à leur utilisation (Goren et al., 2006). Un groupe de protéines riches en résidus d'arginines et sérines, nommées protéines SR, peut être recruté par ces motifs et faciliter l'assemblage des composantes du « spliceosome » aux sites d'épissage 5' et 3' (figure 3A) (Graveley, 2000). Lors de l'utilisation de sites d'épissage non optimaux, la liaison de la snRNP U1 au site d'épissage 5' peut promouvoir la reconnaissance d'un site d'épissage 3' situé en amont : ce mécanisme permettrait de définir l'exon (Robberson et al., 1990). Ce sont les protéines SR qui permettraient la communication entre ces deux sites en interagissant avec la composante protéique de la snRNP U1 (U1-70kDa) et la molécule U2AF35 (figure 3B). De la même façon, la définition de l'intron pourrait aussi être accomplie grâce à l'interaction simultanée des protéines SR avec U1-70kDa et U2AF35 le long de l'intron, contribuant à augmenter la communication entre les sites d'épissage 5' et 3' (Wu et Maniatis, 1993) (figure 3C).

Le contrôle de l'épissage alternatif

La maturation des ARN pré-messagers peut donner lieu à plusieurs patrons d'épissage alternatif où certains segments d'ARN sont gardés ou enlevés (figure 4). La modulation des choix des sites d'épissage alternatif est un processus soumis au contrôle

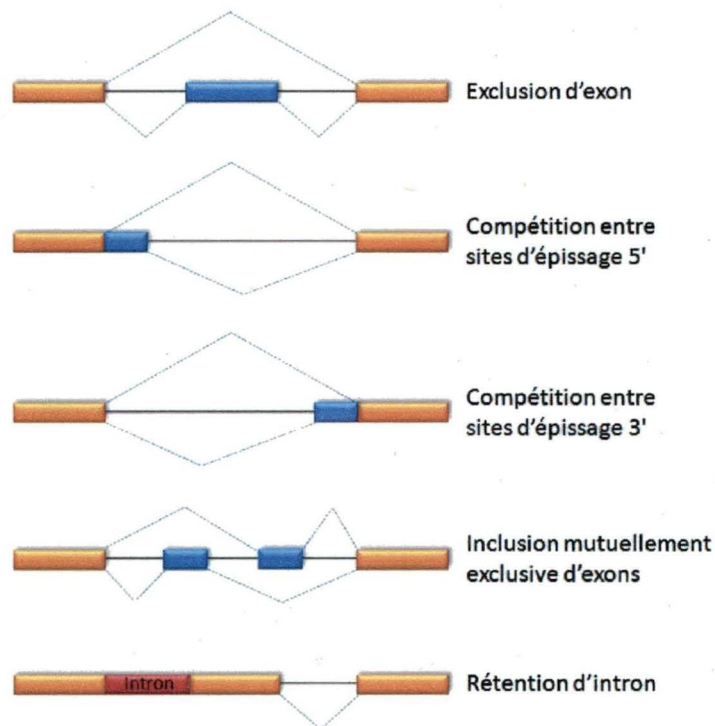


Figure 4. L'épissage alternatif. Lors de la maturation des ARN pré-messagers, les exons constitutifs (rectangles jaunes) sont ligués entre eux, alors que des portions du transcrit peuvent être incluses ou exclues (rectangles bleus). Les modes d'épissage alternatif comprennent l'exclusion d'exons, l'inclusion mutuellement exclusive d'exons, la compétition entre sites d'épissage 5' ou 3' et la rétention d'intron (rectangle rouge).

rigoureux de plusieurs séquences régulatrices. Des régions dites activatrices peuvent être situées dans les exons ou les introns et permettent d'augmenter la liaison des composantes du « spliceosome » aux sites d'épissage alternatif. D'un autre côté, des séquences inhibitrices introniques ou exoniques peuvent nuire à la sélection de sites d'épissage alternatif.

L'identification d'une panoplie de séquences régulatrices et leur localisation dans le génome a permis de mieux comprendre comment s'effectue la sélection des sites d'épissage alternatif et de mieux prédire celle-ci. L'étude d'événements d'épissage alternatif de gènes en particulier, permettant de disséquer les motifs impliqués dans des

unités d'épissage alternatif spécifiques, est une des stratégies utilisées afin de trouver et localiser des éléments régulateurs. D'un autre côté, des études plus globales, mettant à profit des analyses bioinformatiques, ont permis d'identifier des régions du génome qui sont enrichies de séquences régulatrices. Une première approche consiste en la comparaison de gènes homologues entre organismes et à l'identification de régions hautement conservées. Plusieurs études de ce genre ont été effectuées chez la souris et l'humain et ont révélé que les séquences homologues situées dans les exons sont mieux conservées que celles contenues dans les introns : 85% et 69 % d'identité respectivement (Sorek et Ast, 2003). Cependant, les introns bordant les exons alternatifs sont beaucoup plus conservés (88% pour la région en amont et 80% pour celle située en aval) que ceux situés de chaque côté des exons constitutifs (Sorek et Ast, 2003). Toujours en comparant les séquences homologues entre ces deux organismes, on remarque que les exons constitutifs sont moins bien conservés comparativement aux exons alternatifs : 88% et 94% respectivement (Sorek *et al.*, 2004; Sugnet *et al.*, 2004). Compte tenu du fait que des séquences ont été préservées durant l'évolution et que parmi celles-ci, certaines se retrouvent plus fréquemment à proximité de sites d'épissage alternatif, il est donc possible que certains de ces motifs soient impliqués dans la sélection de ces sites. De plus, l'observation que certaines régions qui ne sont pas exprimées peuvent être malgré tout très bien conservées pourrait impliquer un rôle de celles-ci en tant qu'éléments régulateurs de l'épissage alternatif. C'est dans cette optique que le groupe du Dr Ast (Goren et al., 2006) a entrepris une comparaison d'exons homologues entre la souris et l'humain et a identifié plusieurs séquences régulatrices conservées et fréquemment retrouvées dans des exons alternatifs. La suite logique de ce type d'expérience est de confirmer la fonctionnalité de ces motifs par l'emploi d'un système d'épissage hétérologue. Lorsque

ces séquences sont introduites dans un autre contexte d'épissage, elles peuvent activer ou inhiber l'inclusion d'un exon alternatif, démontrant l'activité des éléments individuels.

La prévalence de certains motifs à proximité de certains sites d'épissage peut aussi être étudiée en vérifiant quels types de séquences sont enrichis dans des régions données, non pas grâce à la comparaison de séquences homologues entre différents organismes, mais à partir d'une banque de séquences génomiques humaines. De cette façon, il est possible de vérifier parmi toutes les séquences humaines disponibles qui contiennent des sites d'épissage alternatifs, quels sont les motifs retrouvés plus fréquemment dans ces régions précises. Cette méthode a été utilisée et a confirmé la présence de motifs enrichis près des sites d'épissage alternatif : dix classes d'éléments positifs, ayant tous une activité dans un système d'épissage hétérologue se retrouvent à cet endroit (Fairbrother et al., 2002). Cette étude appuie donc le fait que les sites d'épissage alternatif se distinguent des sites constitutifs par les séquences qui les entourent.

Un rapporteur fluorescent a aussi été utilisé afin d'effectuer un criblage de séquences aléatoires permettant d'enrichir des motifs causant l'exclusion d'un exon alternatif. Cette étude a permis de générer une liste de 133 candidats ayant des activités inhibitrices (Wang et al., 2004). Afin de compléter ces travaux, une analyse statistique a été effectuée, permettant de déterminer si ces séquences sont retrouvées plus fréquemment à certaines positions dans le génome. Cette étude a révélé que ces éléments régulateurs sont retrouvés plus fréquemment près de sites d'épissage alternatif (Wang *et al.*, 2004; Wang *et al.*, 2006), appuyant ainsi le rôle potentiel de ces candidats dans le contrôle de l'épissage alternatif.

L'identification d'un nombre imposant de séquences régulatrices et la façon dont celles-ci sont disposées aidera certes à la prédiction d'événements d'épissage alternatif. Cependant, un aspect important doit être considéré dans la description de ces motifs : l'activité de ceux-ci peut varier considérablement lorsque leur positionnement change dans une unité d'épissage. Par exemple, lorsqu'un motif dit « activateur » est inséré à différents endroits dans un mini-gène contenant l'exon alternatif 8 du gène *ADAR2*, le profil d'épissage alternatif varie de façon marquée (figure 5). En effet, le pourcentage d'inclusion de cet exon est sujet à plusieurs variations lorsque le positionnement de l'élément change : ce motif peut avoir un effet activateur lorsqu'inséré à 32 et 41 nucléotides du site d'épissage 3', alors que la nature de cette même séquence peut être négative lorsque celle-ci est située aux autres positions (Goren et al., 2006). Cette expérience nous suggère donc que la définition d'une séquence régulatrice ne peut pas être prédite que par sa séquence et que l'environnement bordant ce motif doit être considéré dans la définition d'un motif.

La façon dont les séquences régulatrices peuvent moduler le choix des sites d'épissage alternatif peut passer par le recrutement de facteurs spécifiques. Les protéines SR ont initialement été décrites comme étant des régulateurs positifs dans la sélection des sites d'épissage alternatif. Cette classe de facteurs est caractérisée par un ou deux domaines de liaison à l'ARN, en position N-terminale, et une région riche en résidus arginines et sérines, d'une longueur variable, en position C-terminale, qui permet des interactions avec d'autres protéines (Graveley, 2000). La présence de motifs de liaison pour les protéines SR, près des sites d'épissage alternatif, peut stimuler la liaison de la snRNP U1, snRNP U2 et U2AF (Lavigne *et al.*, 1993; Kohtz *et al.*, 1994; Wang *et al.*,

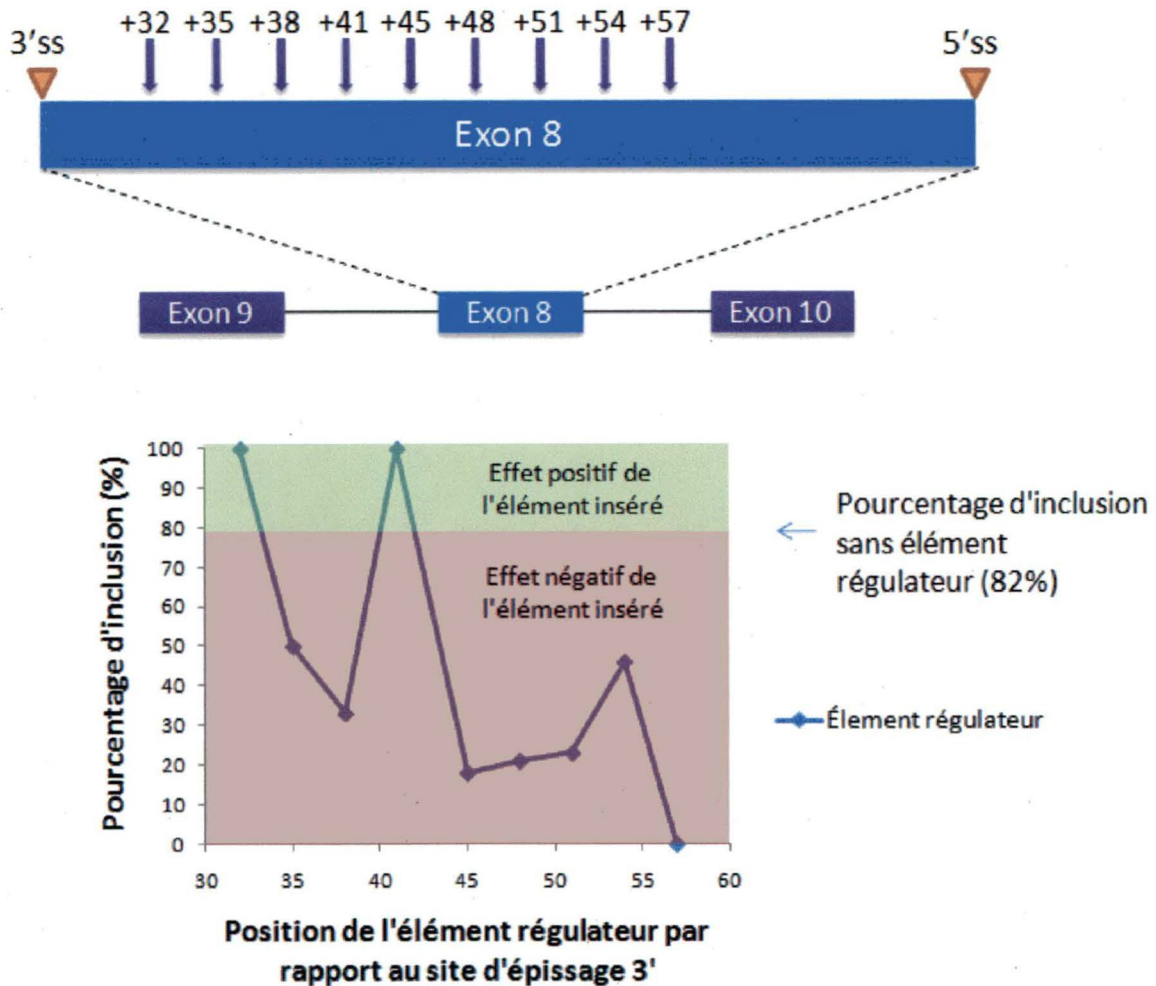


Figure 5. Effet du positionnement d'un élément régulateur. (Partie supérieure de la figure) Insertion d'un élément régulateur, dans un mini-gène contenant l'exon alternatif 8 du gène *ADAR2*, aux positions +32, +35, +38, +41, +45, +48, +51, +54 et +57 par rapport au site d'épissage 3'. (Partie inférieure de la figure) Le pourcentage d'inclusion de l'exon alternatif 8 est représenté sous forme de graphique. Lorsque l'exon 8 ne contient aucun élément régulateur, le pourcentage d'inclusion se situe à 82% : si le motif inséré augmente l'inclusion de l'exon par rapport à cette valeur, celui-ci est considéré comme un activateur (effet positif de l'élément inséré), alors que dans le cas d'une activité inverse, le motif est considéré comme un inhibiteur (effet négatif de l'élément inséré). Données tirées de (Goren et al., 2006).

1995;Zuo et Maniatis, 1996;Graveley *et al.*, 2001). Par contre, cette classe de facteurs peut aussi avoir un effet négatif sur l'épissage : c'est le cas de la protéine ASF/SF2 qui peut inhiber la liaison de la snRNP U2 dans l'unité d'épissage L1 d'adénovirus (Kanopka et al., 1996).

Un autre groupe de protéines, nommé hnRNP, est aussi impliqué dans la modulation de l'épissage alternatif. Cette classe de régulateurs comprend un assortiment approximatif de 24 protéines distinctes (Pinol-Roma et al., 1988) dont l'activité n'est pas strictement liée à l'épissage alternatif. De plus, la disposition des domaines protéiques qui composent les membres de ce groupe varie beaucoup. La raison de ces grandes différences provient du fait que ces protéines n'ont pas été isolées à partir de leurs propriétés biochimiques, mais bien par leur capacité de se lier aux transcrits naissants. En effet, lors de la transcription des gènes par l'ARN polymérase II, les ARN pré-messagers, nommés hnRNA (« heterogeneous nuclear RNA »), sont liés par une variété de protéines qui forme un immense complexe ribonucléoprotéique nommé hnRNP (« heterogeneous nuclear ribonucleoprotein»). Ce groupe de protéines est caractérisé par son enrichissement dans les particules sédimentées à 40S lors du traitement à la nucléase micrococcale (nucléase S7) des hnRNA. Des expériences d'immunofluorescence, utilisant des anticorps monoclonaux, ont montré que l'association des protéines hnRNP aux ARN pré-messagers ne s'effectue pas de façon aléatoire (Matunis et al., 1993). En effet, plusieurs études ont montré que l'interaction des différentes protéines hnRNP à l'ARN s'effectue grâce à des motifs de reconnaissance qui guide leur association (Martinez-Contreras et al., 2007).

Parmi ce groupe de protéines est retrouvé le régulateur hnRNP H qui est impliqué dans la modulation de plusieurs unités d'épissage alternatif et son mode de régulation peut

varier beaucoup d'un gène à l'autre. Dans une première situation, l'action de hnRNP H peut s'effectuer en inhibant l'effet positif d'une autre protéine. Par exemple, chez le gène *α -tropomyosine*, hnRNP H a un impact négatif sur l'inclusion de l'exon alternatif 2 en restreignant la liaison du facteur positif 9G8 (Crawford et Patton, 2006). Son action inhibitrice peut aussi passer par l'inactivation du site d'épissage 3' en agissant directement sur la liaison de la composante U2AF35 : c'est le cas de l'exon 2 du gène *tat* chez le virus HIV-1 (Jacquenet et al., 2001). Dans d'autres situations, hnRNP H peut agir positivement dans l'inclusion d'un exon alternatif. On retrouvera ce facteur dans un complexe qui active l'inclusion de l'exon N1 du gène *c-src* dans les cellules neuronales (Chou et al., 1999). Son action activatrice ou inhibitrice peut dépendre de sa position par rapport aux sites d'épissage. Lorsqu'un site de liaison pour hnRNP H est situé à 66 nucléotides du site d'épissage 5' de l'exon 6D du gène *tev* de HIV-1, ce régulateur peut promouvoir le recrutement de la snRNP U1 (Caputi et Zahler, 2002). D'un autre côté, la liaison de hnRNP H directement au site d'épissage 5' des gènes *NF-1* et *TSH β* entraîne une inhibition de la liaison de la snRNP U1 (Buratti et al., 2004).

Malgré de nombreux exemples de l'effet des protéines hnRNP sur des événements d'épissage alternatif, leur spectre et spécificité d'action est encore peu compris. Une étude récente, dans laquelle 56 événements d'épissage alternatif de gènes apoptotiques ont été vérifiés, lorsque la concentration cellulaire de 14 protéines hnRNP a été diminuée, a permis d'observer que l'identité des gènes régulés par ces régulateurs pouvait varier selon le type cellulaire (Venables et al., 2008b). Une unité d'épissage alternatif spécifique est bien souvent sous le contrôle de plusieurs éléments régulateurs mettant en œuvre une variété de facteurs positifs et négatifs : la balance de tous ces facteurs influencera la

décision finale. L'implication du facteur négatif hnRNP A1 et des régulateurs positifs de la famille des protéines SR nous illustre bien cette situation où des décisions d'épissage alternatif peuvent passer par un mécanisme de compétition entre ces molécules antagonistes. Par exemple, la liaison de hnRNP A1 à l'exon 2 du gène *tat* chez HIV-1 peut bloquer la liaison d'une protéine SR (SC35) et inhiber son effet positif sur la reconnaissance du site d'épissage 3' (Zahler et al., 2004). L'épissage alternatif de l'exon 6B du gène de la β -tropomyosine suit aussi cette règle : un site de liaison pour le facteur négatif hnRNP A1 chevauche ceux des régulateurs positifs ASF/SF2 et SC35 dans le contrôle de l'inclusion de cet exon (Expert-Bezancon et al., 2004). Des cas beaucoup plus complexes impliquant la participation de plusieurs facteurs ont aussi été recensés. Le contrôle de la sélection de deux sites d'épissage 5' en compétition de l'exon 2 du gène *Bcl-x* nous démontre bien ces propos : ceux-ci sont sous la gouverne de sept facteurs en plus de l'implication de la voie de signalisation de la protéine kinase C (Garneau *et al.*, 2005;Massiello *et al.*, 2006;Paronetto *et al.*, 2007;Revil *et al.*, 2007;Cloutier *et al.*, 2008)(figure 6).

Puisque plusieurs facteurs peuvent avoir un impact sur la sélection des sites d'épissage d'une même unité, la régulation des niveaux relatifs de ces protéines est donc un aspect primordial. En plus du contrôle de la transcription des régulateurs et de l'implication des micro-ARN dans la modulation de la concentration de ces protéines cellulaires (Boutz *et al.*, 2007;Makeyev *et al.*, 2007), l'épissage alternatif couplé au mécanisme NMD (« Nonsense-Mediated Decay ») joue aussi un rôle primordial. En effet, l'épissage alternatif peut générer des transcrits, contenant des codons stop prématurés, qui seront dégradés par la machinerie du NMD. L'autorégulation du facteur d'épissage

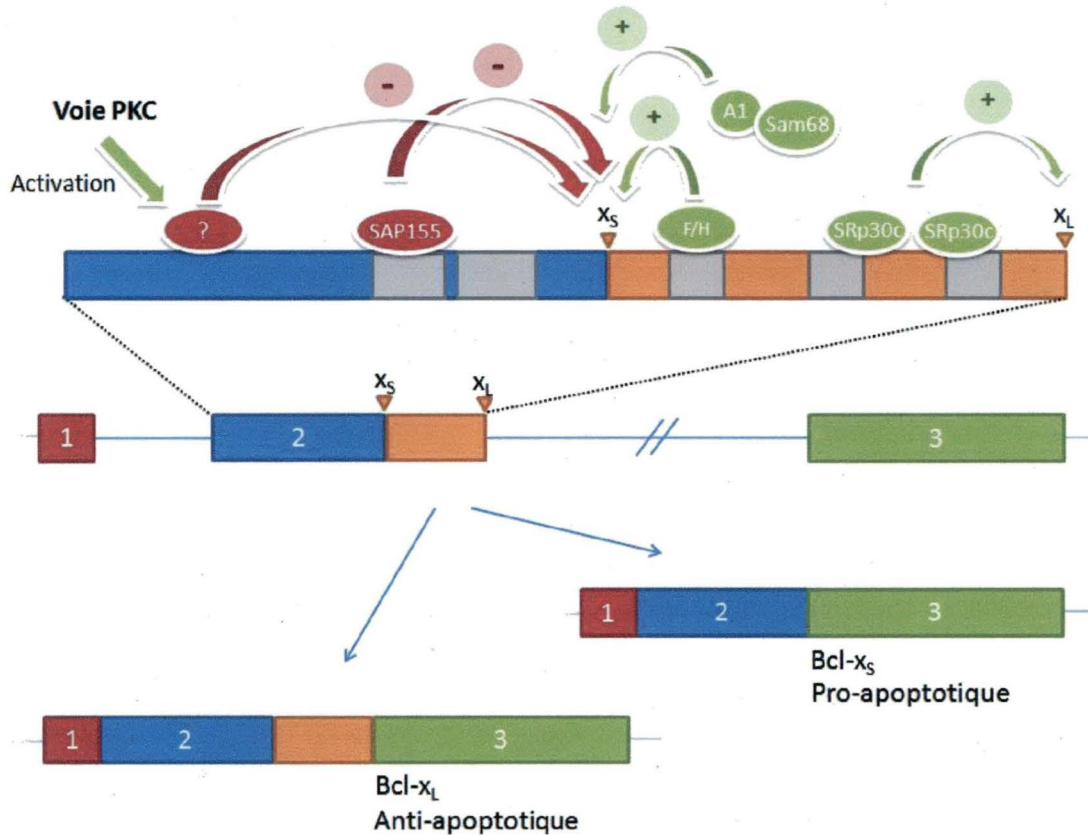


Figure 6. Régulation de l'épissage alternatif de Bcl-x. L'épissage alternatif du gène *Bcl-x* peut générer un isoforme pro-apoptotique (Bcl- x_S) ou anti-apoptotique (Bcl- x_L) via la sélection de sites d'épissage 5' en compétition. La protéine SRp30c peut activer l'utilisation du site X_L , alors que les facteurs hnRNP H, hnRNP F, hnRNP A1 et Sam68 ont un effet positif sur la sélection du site X_S . Le régulateur SAP155 peut inhiber l'utilisation du site X_S , de même qu'un autre facteur encore inconnu qui est sous le contrôle de la protéine kinase C. Tiré de (Fisette *et al.*, 2009).

hnRNP I (PTB) illustre bien cette situation : lorsque cette protéine se retrouve en quantité élevée dans une cellule, celle-ci peut lier son propre ARN pré-messager, induire l'exclusion d'un exon alternatif et mener à un transcrit qui contient un codon stop prématuré (Wollerton *et al.*, 2004). L'homéostasie de hnRNP L peut aussi être régulée de la même façon. Ce régulateur peut activer l'inclusion d'un exon alternatif (contenant un codon stop prématuré) de son propre ARN pré-messager et générer un ARN messager qui sera ensuite dégradé (Rossbach *et al.*, 2009). Une étude dans laquelle la voie du NMD a

été bloquée montre que le profil d'épissage alternatif de plusieurs régulateurs positifs et négatifs est affecté et les isoformes générés contiennent une fréquence élevée de codons stop prématurés (Ni et al., 2007). Étant donné la sensibilité de ces ARN messagers à cette voie de contrôle, ceci suggère que le contrôle des niveaux cellulaires de plusieurs régulateurs pourrait passer par un couplage entre l'épissage alternatif et le NMD.

En plus de la participation des protéines SR et hnRNP, une panoplie d'autres régulateurs sont impliqués dans le contrôle de l'épissage alternatif. On note entre autre que certains facteurs sont exprimés spécifiquement dans quelques tissus ce qui engendre des patrons d'épissage alternatif uniques à ces régions. Par exemple, la protéine FOX2 qui est exprimée exclusivement dans les cellules musculaires, nerveuses et dans les cellules souches embryonnaires humaines a été caractérisée comme un régulateur d'épissage alternatif (Yeo et al., 2009). Plus spécifiquement, dans les cellules souches embryonnaires, le facteur FOX2 a un patron de régulation unique : lorsque cette protéine se lie en amont d'un exon alternatif, elle a tendance à l'inhiber, alors que sa présence en aval active son inclusion (Yeo et al., 2009). Lors de la différenciation des cellules embryonnaires en neurones, les cibles de régulation sont altérées, suggérant l'expression de nouveaux facteurs qui agiront sur les mêmes unités d'épissage que FOX2. D'autres facteurs tels que Nova-1 et Nova-2, exprimés exclusivement dans les neurones, peuvent influencer des décisions d'épissage et le positionnement de leurs sites de liaison déterminera leur effet activateur ou répresseur dans l'épissage alternatif. Par exemple, lorsque des motifs de liaison pour les protéines Nova sont retrouvés dans un exon alternatif, le recrutement de ces régulateurs peut bloquer la liaison de la snRNP U1. Au

contraire, lorsque ces molécules lient une région intronique, en aval d'un exon alternatif, elles peuvent promouvoir l'assemblage du « spliceosome » (Ule et al., 2006).

Mis à part le recrutement de facteurs aux séquences régulatrices, la formation de structures secondaires peut aussi moduler le choix des sites d'épissage alternatif. Par exemple, l'épissage alternatif de l'exon 10 du gène codant pour la protéine Tau (protéine associée aux microtubules) est sous le contrôle d'une tige boucle au site d'épissage 5' qui limite la liaison de la snRNP U1. Or, des mutations dans cette région qui abolissent la complémentarité de la structure rendent plus accessibles le site d'épissage 5' et augmentent l'inclusion de l'exon 10 : cette situation va créer un déséquilibre des ratios d'isoformes du gène *Tau* et être associée à des désordres neurologiques importants (Liu et Gong, 2008). Il existe une situation beaucoup plus complexe chez le gène *DSCAM* (« down syndrome cell adhesion molecule ») de la drosophile. En tenant compte que l'ARN pré-messager produit contient 95 exons alternatifs, il y aurait théoriquement 38 016 isoformes qui pourraient être générés. Une section de ce gène comprend un exon constitutif qui doit être juxtaposé à un seul des 48 exons alternatifs disponibles dans cette région. Afin d'accomplir cette tâche, chaque exon alternatif contient, en amont, une séquence qui est complémentaire à une région située en aval de l'exon constitutif. L'appariement de ces deux séquences permet le rapprochement d'un exon alternatif avec l'exon constitutif : cette juxtaposition d'exons serait le mécanisme central permettant la sélection d'un seul exon alternatif (Graveley, 2005).

Les protéines hnRNP H et hnRNP F

Le facteur hnRNP F fait partie de la même famille de protéines que hnRNP H. Une comparaison des résidus d'acides aminés entre ces deux molécules révèle une identité de 78% (Honore et al., 1995). La majorité des protéines hnRNP contient des motifs de liaison à l'ARN appelés RRM (« RNA recognition motif ») (Martinez-Contreras et al., 2007). Les molécules hnRNP H et hnRNP F possèdent, en position N-terminale, deux domaines peu apparentés au RRM classique qui permettent tout de même la liaison à l'ARN : ceux-ci sont appelés qRRM 1 et qRRM 2 (quasi-RRM) (Honore et al., 1995). En position C-terminale de ces protéines sont retrouvés deux domaines riches en résidus glycines (GRD) entrecoupés d'un troisième qRRM 3 (Honore et al., 1995) (figure 7).

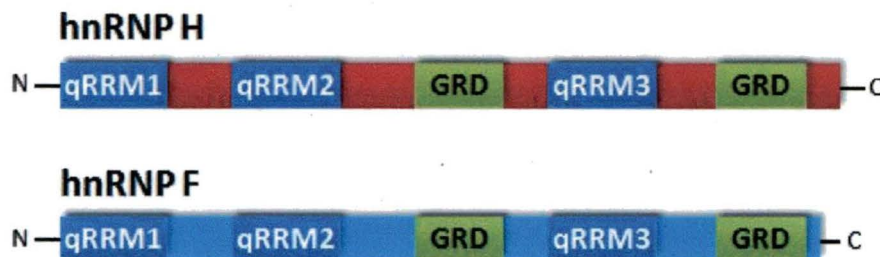


Figure 7. Domaines protéiques de hnRNP F et hnRNP H. Les protéines hnRNP H et hnRNP F contiennent deux quasi-RRM en position N-terminale, ainsi que deux domaines riches en résidus glycines interrompus par un troisième quasi-RRM.

Les études initiales ont établi que les protéines hnRNP F et hnRNP H ont de l'affinité pour des séquences poly-guanosines (Swanson et Dreyfuss, 1988). Plus spécifiquement, une répétition minimale de trois guanosines, qui doit être suivie d'une adénosine, est requise pour la liaison de ces protéines à l'ARN (Caputi et Zahler, 2001). Dans certaines situations, une répétition de plusieurs éléments riches en guanosines

augmente l'affinité de hnRNP H et hnRNP F à l'ARN. De plus, une augmentation du nombre de guanosines consécutives peut aussi aider à une meilleure liaison (Arhin et al., 2002). Il a aussi été observé qu'il peut y avoir une coopération dans la liaison des protéines hnRNP H et hnRNP F aux séquences riches en guanosines (Alkan et al., 2006). Cependant, la prédiction précise des sites de haute affinité pour ces protéines reste ardue, car la disposition des éléments riches en guanosines et l'importance des séquences adjacentes à ces régions sont des aspects peu compris.

De façon générale, ces deux protéines sont exprimées dans la majorité des tissus, mais les niveaux d'expression peuvent différer. On note par exemple que le régulateur hnRNP F est plus fortement exprimé dans les cellules épithéliales de la prostate alors que le facteur hnRNP H est plus prédominant dans les cellules de la peau (Honore et al., 2004). Dans certaines cellules, comme celles de l'épithélium gastrique, aucune expression de ces protéines n'a pu être détectée (Honore et al., 2004). En ce qui a trait à la localisation cellulaire, la protéine hnRNP F est à la fois retrouvée au noyau et au cytoplasme, alors que la molécule hnRNP H est principalement localisée au noyau (Honore et al., 2004).

Dans un autre ordre d'idée, la protéine hnRNP F est aussi impliquée dans des décisions d'épissage alternatif et est quelques fois retrouvé dans les mêmes complexes régulateurs que hnRNP H. C'est le cas du contrôle de l'exon alternatif N1 du gène *c-src*, où ces deux protéines font partie d'un complexe activateur (Min *et al.*, 1995; Chou *et al.*, 1999) et chez le gène *Bcl-x* où celles-ci augmentent l'utilisation d'un site d'épissage 5' distal (Garneau et al., 2005). L'épissage alternatif du gène *α -tropomyosine* implique aussi la participation des protéines hnRNP H et hnRNP F. En effet, celles-ci jouent un rôle

important, puisqu'elles vont inactiver l'inclusion d'un exon alternatif contenant un codon stop prématuré (Coles et al., 2009). La liaison du facteur FOX2 à proximité des molécules hnRNP H et hnRNP F permet d'accroître leur affinité à une séquence régulatrice située dans l'exon alternatif IIIc du gène *FGFR2* (« Fibroblast Growth Factor Receptor 2 »). Étant donné que ces deux protéines ont un effet négatif sur l'inclusion de cet exon et qu'elles sont en compétition avec le facteur positif ASF/SF2 pour la liaison à une même région exonique, la présence de FOX2 augmenterait ainsi l'effet inhibiteur de hnRNP H et hnRNP F (Mauger et al., 2008). Mis à part leur implication dans l'épissage alternatif, notons entre autre leur rôle antagoniste dans la maturation de l'extrémité 3' des ARN pré-messagers. hnRNP H peut activer le processus de polyadénylation en se liant à des éléments riches en guanosines situés en amont du signal de polyadénylation (Arhin et al., 2002), alors que hnRNP F inactive la réaction de clivage de l'extrémité 3' en bloquant la liaison du facteur responsable de la coupure CstF-64 (« cleavage stimulation factor ») (Veraldi et al., 2001).

Épissage alternatif de l'ARN pré-messager de hnRNP A1

Afin d'étudier les mécanismes impliqués dans la régulation des sites d'épissage, notre laboratoire a décidé d'élucider l'épissage alternatif du gène *hnRNP A1*. Deux isoformes peuvent être produits à partir de ce gène : un transcrit qui inclut ou exclut l'exon 7B pour générer un ARN messager codant pour la protéine hnRNP A1B ou hnRNP A1 respectivement. Dans la plupart des types cellulaires, l'isoforme hnRNP A1 est produit majoritairement (Buvoli *et al.*, 1990a; Hanamura *et al.*, 1998). La dissection des éléments régulateurs a débuté par une comparaison entre la souris et l'humain des introns bordant l'exon alternatif 7B. Étant donné que ces régions ne sont pas exprimées, les

séquences conservées entre ces deux organismes pourraient donc être impliquées dans le contrôle de l'exclusion de cet exon. À titre d'exemple, l'intron situé entre les exons constitutifs 6 et 7 est conservé à environ 25% alors que les séquences introniques situées de chaque côté de l'exon 7B le sont entre 60 et 70 % (Chabot et al., 2003). Parmi les dix régions repérées comme étant très conservées, quatre éléments introniques impliqués dans le contrôle de l'épissage de l'exon alternatif 7B ont été identifiés. Un premier motif, situé en aval de l'exon 7B, est responsable de la formation d'une structure secondaire impliquant l'appariement de cette région au site d'épissage 5' de cet exon : cela a pour effet de bloquer l'accès à la snRNP U1 (Blanchette et Chabot, 1997). En amont de l'exon 8, se situe une région intronique ayant la capacité de recruter SRp30c et d'inhiber l'utilisation du site d'épissage 3' de cet exon (Simard et Chabot, 2000; Simard et Chabot, 2002). Ce blocage déplacerait l'utilisation du site d'épissage 3' vers celui de l'exon 7B, augmentant ainsi son inclusion. Cependant, la protéine PTB peut être recrutée à la même région, réduisant la liaison de SRp30c et son action inhibitrice (Paradis et al., 2007). Finalement, deux régions situées de part et d'autre de l'exon 7B ont été identifiées comme étant impliquées dans son exclusion (Chabot *et al.*, 1997; Blanchette et Chabot, 1999; Hutchison *et al.*, 2002). Ces séquences permettent le recrutement de la protéine hnRNP A1 et la formation d'homodimères entre ces molécules qui cause un réarrangement dans la structure de l'ARN pré-messager (Chabot *et al.*, 1997; Blanchette et Chabot, 1999; Martinez-Contreras *et al.*, 2006). Cette modulation de l'ARN engendrerait le rapprochement du site d'épissage 5' de l'exon 7 et 3' de l'exon 8 tout en masquant ceux de l'exon 7B (figure 8). Plusieurs observations supportent ce modèle. Premièrement, une interaction entre protéines hnRNP A1 est possible et la région riche en résidus glycines est requise pour cette interaction (Cartegni et al., 1996). De plus, l'assemblage d'un complexe

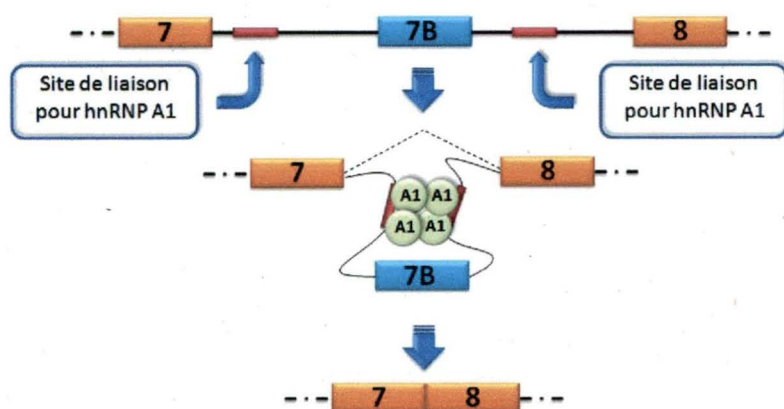


Figure 8. Modèle de changement de conformation de l'ARN pré-messager de hnRNP A1 dans l'exclusion de l'exon 7B. Des interactions entre protéines hnRNP A1, liées de chaque côté de l'exon 7B, permettent le rapprochement du site d'épissage 5' de l'exon 7 et 3' de l'exon 8.

entre molécules hnRNP A1 serait stabilisé par la liaison des partenaires à l'ARN (Blanchette et Chabot, 1999). Dans ce mode d'action proposé, qui implique un changement de conformation de l'ARN pré-messager, il a été démontré que la liaison des snRNP U1 aux sites d'épissage 5' n'est pas affectée : ce qui est en accord avec le modèle qui prédit un rapprochement des sites distants sans que ceux-ci ne soient affectés (Chabot et al., 1997).

L'utilisation d'une unité d'épissage *in vitro* constituée de deux sites d'épissage 5' (distal et proximal) en compétition pour un seul site d'épissage 3' nous a permis de raffiner notre compréhension du modèle dans lequel des molécules hnRNP A1 peuvent interagir et rapprocher des sites distants. Tout d'abord, la présence de sites de haute affinité pour la protéine hnRNP A1, de chaque côté du site d'épissage 5' proximal, active la sélection du site d'épissage 5' distal (Blanchette et Chabot, 1999). L'activité de hnRNP A1 sur cette unité d'épissage est dépendante de son domaine riche en résidus glycines : un mutant de hnRNP A1, contenant seulement les deux premiers RRM, peut lier l'ARN,

mais agit comme dominant négatif (Blanchette et Chabot, 1999). L'insertion de séquences en amont de chaque site de liaison pour hnRNP A1 (permettant de les éloigner des sites d'épissage 5') n'affecte pas le déplacement vers le site d'épissage 5' distal (Nasim et al., 2002). Cette dernière expérience suggère que l'hypothèse d'une activation directe du site d'épissage 5' distal ou d'une répression du site d'épissage 5' proximal est peut envisageable. Il est aussi possible de reproduire l'effet observé en remplaçant les sites de liaison pour hnRNP A1 par des séquences complémentaires, appuyant ainsi le modèle de rapprochement des sites distants dans le remodelage de la structure de l'ARN pré-messager (Nasim et al., 2002).

L'objectif global de notre étude consistait à mieux comprendre comment les protéines hnRNP H peuvent moduler l'épissage alternatif. Plus précisément, nos travaux étaient basés sur le modèle proposé avec hnRNP A1 (décrit plus haut) et ont été initiés dans le but de vérifier l'hypothèse suivante : la protéine hnRNP H peut moduler l'épissage alternatif de la même manière que hnRNP A1. Nos expériences d'épissage *in vitro* nous ont permis de confirmer cette affirmation. Suite à cela, nous avons démontré que les protéines hnRNP H et hnRNP A1 peuvent coopérer pour moduler la sélection des sites d'épissage 5'. Nos objectifs ont ensuite été dirigés à étudier les déterminants moléculaires importants dans l'activité observée avec hnRNP H. Ces travaux ont été réalisés par l'intermédiaire de protéines hybrides qui nous ont permis d'observer que les protéines hnRNP A1 et hnRNP H requièrent les mêmes domaines dans leur activité d'épissage alternatif : le domaine riche en résidus glycines est un déterminant clé. Nos prémisses concernant l'activité d'épissage des protéines hnRNP H et hnRNP A1 étaient basées sur la formation d'interactions protéines : protéines entre ces molécules. Afin de vérifier cette

dernière proposition, nous avons utilisé la technologie BRET. Nous avons détecté une forte proximité entre ces protéines, dans les cellules vivantes, suggérant une interaction entre ces molécules.

Sommes toutes, nos travaux auront permis de confirmer nos hypothèses. En effet, nos résultats supportent la formation d'homodimères et d'hétérodimères entre les protéines hnRNP A1 et hnRNP H qui permet d'engendrer un changement de conformation de l'ARN pré-messager et de moduler les sites d'épissage alternatif. De façon plus globale, nos données recueillies aideront à mieux comprendre comment les protéines hnRNP A1 et hnRNP H peuvent moduler des décisions d'épissage alternatif et nous guideront à mieux prédire celles-ci.

CHAPITRE 1

*Martinez-Contreras R., *Fisette J-F., *Nasim F., Cordeau M., Chabot B., Intronic binding sites for hnRNP A/B and hnRNP F/H proteins stimulate pre-mRNA splicing. *PLoS Biol*, 2006. 4(2): p. e21. (***ces auteurs ont contribué également**).

RÉSUMÉ

Plusieurs événements d'épissage alternatif impliquent la participation de la protéine hnRNP A1. Nos résultats démontrent que des sites de liaison pour cette molécule peuvent stimuler l'épissage *in vitro* d'un ARN pré-messager contenant un intron artificiellement élargi. Cette activité était observée lorsque les éléments de liaison étaient introduits aux extrémités de l'intron et peut être reproduite en utilisant un oligonucléotide contenant une portion complémentaire à l'intron et une section contenant un site de liaison pour hnRNP A1. Dans un contexte *in vivo*, l'inclusion d'un exon alternatif bordant un intron artificiellement élargi peut être activée grâce à cet oligonucléotide. Des sites de liaison pour hnRNP F/H peuvent aussi stimuler l'épissage d'un ARN pré-messager contenant un intron élargi. De plus, une combinaison de sites de liaison pour hnRNP F/H peut moduler la sélection des sites d'épissage 5'. Le modèle retenu afin d'expliquer ces activités implique une interaction entre molécules hnRNP liées à l'ARN qui permet de rapprocher des sites d'épissage distants. Finalement, la distribution des sites de liaison pour hnRNP F/H et hnRNP A1 révèle une fréquence plus élevée de ceux-ci aux extrémités des introns, supportant ainsi une coopération entre les sites de liaison. Nos travaux suggèrent ainsi que la conformation des ARN pré-messagers peut être modulée par les protéines hnRNP A1 et hnRNP F/H afin de diriger la sélection des sites d'épissage.

J'ai contribué à la réalisation des expériences qui démontrent que hnRNP H peut moduler la sélection des sites d'épissage 5'. Plus précisément, j'ai réalisé les expériences d'épissage *in vitro* montrant que des sites de liaison pour hnRNP F/H peuvent reproduire l'activité des sites de liaison pour hnRNP A1. J'ai aussi réalisé les expériences d'épissage

in vitro montrant que l'activité d'épissage observée est dépendante de la protéine hnRNP H. Finalement, j'ai effectué les expériences de protection à la RNase H montrant que la liaison de la snRNP U1 aux sites d'épissage 5' n'est pas affectée en présence de sites de liaison pour hnRNP H. J'ai contribué à l'écriture de cet article.

Intronic Binding Sites for hnRNP A/B and hnRNP F/H Proteins Stimulate Pre-mRNA Splicing

Rebeca Martinez-Contreras^{1*}, Jean-François Fisette^{1*}, Faiz-ul Hassan Nasim^{1*}, Richard Madden², Mélanie Cordeau¹ and Benoit Chabot^{1,2*}

RNA/RNP Group. ¹ Département de microbiologie et d'infectiologie, ² Centre de Génomique Fonctionnelle de Sherbrooke, Faculté de médecine et des sciences de la santé, Université de Sherbrooke, Sherbrooke, Québec, Canada. J1H 5N4

*Corresponding author: email: Benoit.Chabot@USherbrooke.ca

* These authors contributed equally to this work

Running title: hnRNP A/B and F/H proteins as splicing factors

Keywords: antisense/bifunctional oligonucleotides/hnRNP proteins/introns/RNA conformation

ABSTRACT

hnRNP A/B proteins modulate the alternative splicing of several mammalian and viral pre-mRNAs, and are typically viewed as proteins that enforce the activity of splicing silencers. Here we show that intronic hnRNP A/B binding sites (ABS) can stimulate the *in vitro* splicing of pre-mRNAs containing artificially enlarged introns. Stimulation of *in vitro* splicing could also be obtained by providing intronic ABS *in trans* through the use of antisense oligonucleotides containing a non-hybridizing ABS-carrying tail. ABS-tailed oligonucleotides also improved the *in vivo* inclusion of an alternative exon flanked by an enlarged intron. Notably, binding sites for hnRNP F/H proteins (FBS) replicate the activity of ABS by improving the splicing of an enlarged intron and by modulating 5' splice site selection. One hypothesis formulated to explain these effects is that bound hnRNP proteins self-interact to bring in closer proximity the external pair of splice sites. Consistent with this model, positioning FBS or ABS at both ends of an intron was required to stimulate splicing of some pre-mRNAs. In addition, a computational analysis of the configuration of putative FBS and ABS located at the ends of introns support the view that these motifs have evolved to support cooperative interactions. Our results document a positive role for the hnRNP A/B and hnRNP F/H proteins in generic splicing, and suggest that these proteins may modulate the conformation of mammalian pre-mRNAs.

INTRODUCTION

Exons represent approximately 1% of the human genome and range in size from 1 to 1000 nt, with a mean size for internal exons of 145 nt [1]. In contrast, introns constitute 24% of our genome with sizes ranging from 60 to more than 200 000 nt. The mean size of human introns is more than 3 300 nt and nearly 20% of human introns are larger than 5 Kb [1]. While the efficient and accurate removal of introns is crucial for the production of functional mRNAs, it remains unclear how an intron is defined when splicing junctions are separated by thousands of nucleotides. Although intron size can influence alternative splicing in mammals [2], the mechanisms that enforce the removal of large mammalian introns have not been investigated partly because introns larger than 1 kb are not spliced efficiently in vitro.

Some of the decisions associated with the removal of large introns are likely to be similar to the choices made by the splicing machinery when selecting alternative splice sites. Choosing the appropriate pair of splice sites in alternative splicing units requires the contribution of many types of elements that are recognized by different classes of proteins including SR and hnRNP proteins. hnRNP A1 was the first protein of its class being attributed a function in splice site selection based on its ability to antagonize the activity of the SR protein ASF/SF2 in an in vitro 5' splice site selection assay [3]. hnRNP A/B proteins have now been documented to modulate the alternative splicing of many mammalian and viral pre-mRNAs [for a review, see 4]. In one case, an exonic binding site for hnRNP A1 prevents the interaction of positively-acting SR proteins by a process that apparently involves the nucleation of several A1 molecules [5]. Our previous work on the

alternative splicing of the hnRNP A1 pre-mRNA led to a different model to explain the activity of hnRNP A/B proteins bound to intronic sites. In this case, bound hnRNP A/B proteins would self-interact to loop out and repress internal splice sites [4,6,7]. This postulated change in pre-mRNA conformation would bring in closer proximity the external pair of exons, an event that may also enforce intron definition. Notably, putative binding sites for hnRNP A/B are more abundant in introns than in exons (1.2 site and 0.35 site per 1000 nt, respectively), and are found more frequently near splice junctions [6]. A similar bias in the distribution of GGG motifs has been observed in mammalian introns [8,9,10,11,12,13,14,15,16]. GGG motifs located downstream from a 5' splice site facilitate intron definition [13,17,18,19], and their insertion near the 5' splice site of a Fugu intron stimulates splicing in a mammalian cell line [15]. Because many binding sites for hnRNP A/B and F/H proteins contain G triples [20,21,22,23], hnRNP A/B and hnRNP F/H proteins bound near the ends of an intron may be responsible for the activity of GGG motifs and, hence, may play a generic role in intron definition.

We have examined the role of hnRNP A/B and hnRNP F/H proteins in intron removal by making use of the observation that enlarged introns are spliced poorly in vitro. We show that cis- and trans-acting binding sites for hnRNP A/B and hnRNP F/H proteins (ABS and FBS, respectively) stimulate the splicing of such introns. For some pre-mRNAs, the splicing enhancing activity of ABS or FBS requires that these sites be positioned at both ends of the intron, consistent with the view that hnRNP A/B and F/H proteins act cooperatively to change pre-mRNA conformation. Overall, our results suggest that hnRNP A/B and hnRNP F/H proteins may stimulate splicing by facilitating intron definition.

RESULTS

Binding sites for hnRNP A/B proteins stimulate the in vitro splicing of enlarged introns

To monitor splicing activity, we first used model pre-mRNAs with small introns that are spliced efficiently in vitro (7-Ad and 7B-Ad; Figure 1A) [7]. As determined by the RT-PCR amplification of the resulting mRNA products, a time-course co-incubation of these two pre-mRNAs in a HeLa nuclear extract indicates that they are spliced with similar efficiencies (Figure 1B, lanes 1-6). Insertion of a 1015 nt-long lambda fragment in the intron of each model substrates generated the 7-AdA(-.) and 7B-AdA(-.) derivatives. These larger pre-mRNAs were spliced less efficiently than the small intron versions (Figure 1B, lanes 7-12 and accompanying graph). A splicing efficiency inferior to 1% was estimated by visualizing labeled lariat molecules in a conventional splicing gel (Figure 1C, lane 2). Different lambda inserts of similar lengths had a similar effect on splicing, as seen when the 7-AdB(-.) and 7-AdC(-.) pre-mRNAs were co-incubated in splicing mixtures with the small-intron 7B-Ad pre-mRNA (Figure 1D, lanes 2-6). Thus, increasing intron size with a variety of lambda inserts reduced in vitro splicing efficiency.

To determine whether hnRNP A/B proteins could stimulate the splicing of these enlarged introns, a hnRNP A/B binding site (ABS) corresponding to the CE1a element in the mouse hnRNP A1 pre-mRNA [6,24] was inserted at two positions in the enlarged introns: one ABS was inserted 26 nt downstream from the 5' splice junction, while the second ABS was inserted 88 nt upstream from the 3' splice junction (Figure 1A). The presence of two ABS in the 7-AdA(a.a) and 7B-AdA(a.a) pre-mRNAs stimulated

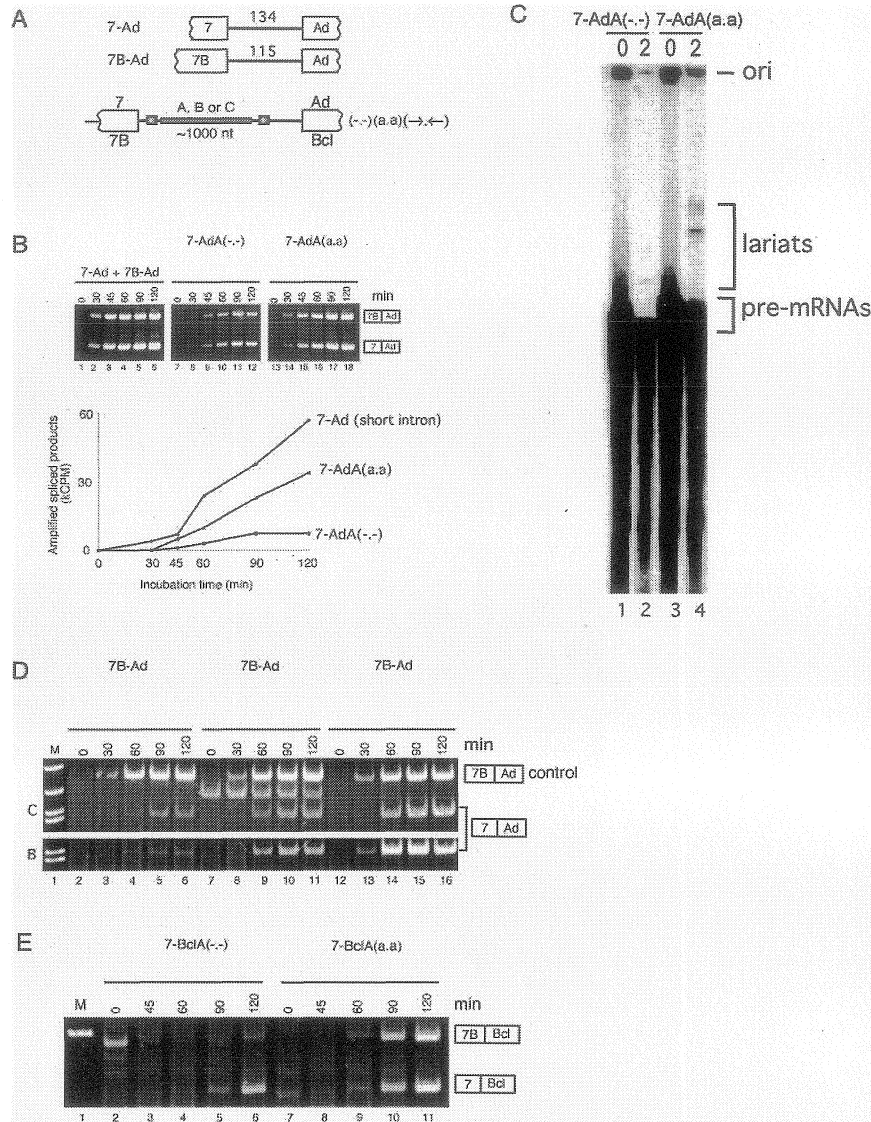


Figure 1. Binding sites for hnRNP A1/A2 stimulate the in vitro removal of enlarged introns. (A) The model pre-mRNAs contain portions of exons 7 or 7B of the hnRNP A1 gene paired with the adenovirus L2 exon. The size of the small introns in 7-Ad and 7B-Ad pre-mRNAs is indicated in nucleotides. The size of lambda inserts A, B and C are respectively 1015, 943 and 1038 nt. The lambda inserts do not contain the sequences $UAGGG^{U/A}$ or $UAGAG^{U/A}$ which correspond to high-affinity binding sites for hnRNP A/B proteins [20,43]. The larger intron substrates contain either exon 7 or exon 7B as first exon, and either the adenovirus L2 or the Bcl-x exon 3 as second exon. When no other elements are inserted, the pre-mRNAs correspond to the (-.) version. The (a.a) versions contain ABS inserted 26 nt downstream of the 5' splice site and 88 nt upstream of the 3' splice site. The (→.←) versions contain inverted repeats at the same positions as ABS. (B) The 7-Ad and 7B-Ad pre-mRNAs were co-incubated for the times indicated (in minutes) in a HeLa nuclear extract (lanes 1-6). Additional mixtures were prepared with

pre-mRNAs carrying lambda insert A lacking or containing ABS (lanes 7-12 and 13-18, respectively). The concentration of each pre-mRNA was 80 pM. Following RNA extraction, the mRNA products from mixtures were amplified by RT-PCR using a common set of primers (reverse primer complementary to the adenovirus exonic sequence and forward primer corresponding to plasmid sequence upstream of exon 7 or exon 7B sequences). The graph displays the abundance of amplified splicing product at different times for 7-Ad and the different 7-AdA pre-mRNAs. The RT-PCR assay shown here and in other figures was performed in conditions that displayed a linear relationship between the amounts of input RNA and amplified products over a large range of input RNA concentrations (from 10-fold less to at least 6-fold more than the amounts used in the assays (data not shown)). (C) Splicing reactions were set using ³²P-labeled pre-mRNAs and incubated for 0 or 2 hours in HeLa nuclear extracts. Total RNA was extracted and the splicing products were fractionated on a 5% acrylamide/8 M urea gel. The position of the lariat products is indicated. (D) Each of the 7-Ad pre-mRNAs carrying lambda inserts B or C (7-AdB or 7-AdC; 80 pM) was co-incubated with the small-intron 7B-Ad pre-mRNA (8 pM). Versions lacking (-) or containing ABS (a.a), as well as carrying inverted repeats (→.←) were used. Following incubation for different times, spliced products were amplified by RT-PCR using a common set of primers. The co-incubated small-intron control is only shown for the 7-AdC pre-mRNA mixture. M = molecular weight markers. (E) Large-intron pre-mRNAs 7-BclA and 7B-BclA (80 pM each) lacking (-) or containing ABS (a.a) were co-incubated for the indicated times in a HeLa extract. RT-PCR was performed as above except that a Bcl-x reverse primer was used. The band amplified at t = 0 (lane 2) is artifactual and does not co-migrate with the 7B/Bcl splicing product. M = molecular weight markers.

splicing approximately 4-fold (Figure 1B, lanes 13-18 and accompanying graph).

Likewise, when lariat splicing products were detected in a conventional splicing gel using ³²P-labeled transcripts, the presence of ABS stimulated in vitro splicing 3-fold (Figure 1C, lane 4). Splicing stimulation was also observed when ABS were inserted in the 7-AdB and 7-AdC pre-mRNAs (Figure 1D, lanes 7-11). Stimulation was estimated to be in the order of 3 to 5-fold based on amplification reactions performed in the presence of ³²P-dCTP (not shown). Replacing ABS with 20 nt-long inverted repeats also stimulated large-intron splicing (Figure 1D, lanes 12-16), suggesting that looping out intron sequences can improve the in vitro splicing efficiency of enlarged introns. We also tested the impact of ABS on the splicing of large-intron pre-mRNAs carrying the 3' splice site of Bcl-x exon 3

(7-BclA and 7B-BclA). The presence of ABS also stimulated the production of amplicons corresponding to mRNA products (Figure 1E, compare the intensities of the 7/Bcl and 7B/Bcl products in lanes 2-6 with those in lanes 7-11).

To insure equivalent recovery and loading of the various samples in future experiments, we relied on systematically co-incubating each test pre-mRNA with a control small-intron pre-mRNA in splicing mixtures. mRNAs products derived from both substrates could then be amplified simultaneously by RT-PCR using the same primer set. In these conditions, we confirmed that the stimulation offered by ABS was observed at different ratios of test/control pre-mRNAs (Figure S1). To confirm the participation of hnRNP A/B proteins in splicing stimulation, we added to a HeLa nuclear extract increasing amounts of a DNA oligonucleotide (TS10) carrying high-affinity binding sites for A1 and A2 [apparent K_d below 5 nM, data not shown and see 25,26]. We have shown

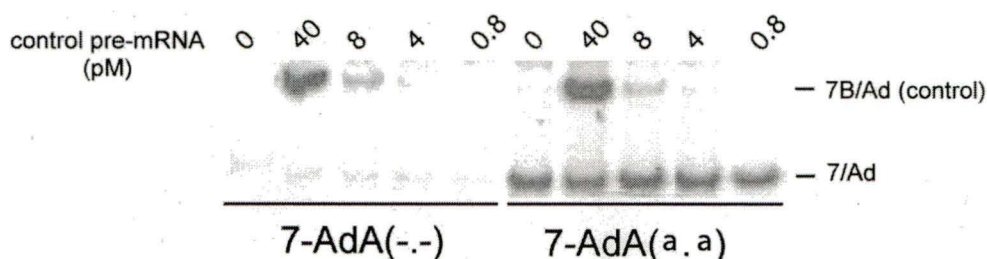


Figure S1. RT-PCR assay of splicing mixtures incubated with different ratios of test and control pre-mRNAs. Eighty pM of test 7-AdA(-.-) or 7-AdA(+) pre-mRNA was mixed with various amounts of control 7B-Ad pre-mRNA (from 0 to 40 pM). The mixtures were incubated in HeLa nuclear extracts for two hours at 30°C. The RT-PCR assay was carried out with a single pair of primers that amplify mRNA products derived from all pre-mRNAs. The stimulation provided by ABS can be observed at all ratios of test/control pre-mRNAs.

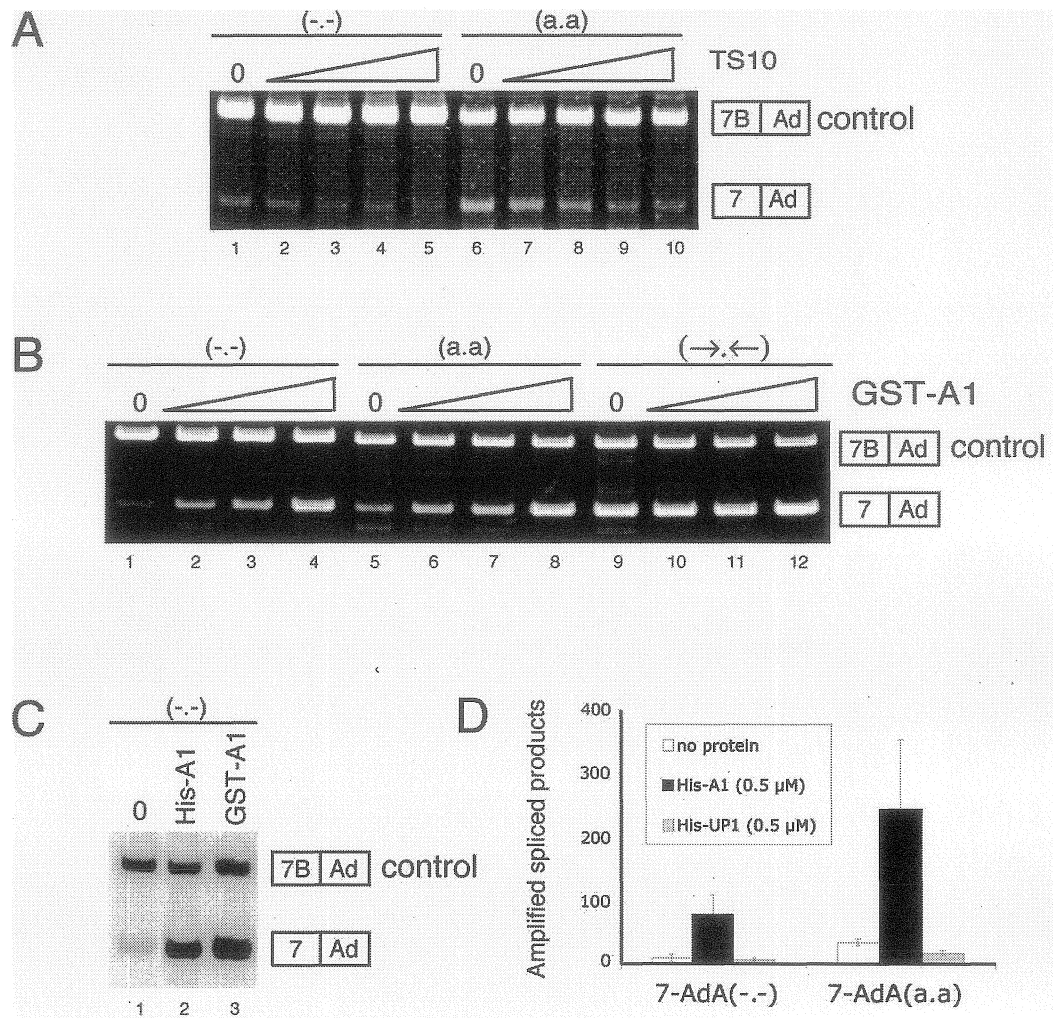


Figure 2. The hnRNP A1 protein stimulates large-intron splicing. (A) Sequestering hnRNP A/B proteins affects large-intron splicing. The large-intron substrates 7-AdB lacking ABS (-.-) or containing ABS (a.a) were co-incubated with the small-intron 7B-Ad pre-mRNA (80 pM for the 7-AdB substrates and 16 pM for the control 7B-Ad pre-mRNA) in a HeLa extract for 90 min in the presence of increasing amounts of the telomeric oligonucleotide TS10 (0, 80, 160, 320, 640 nM, respectively). (B) Splicing mixtures were incubated with increasing amounts of recombinant GST-A1 protein (0, 0.8, 1.6 and 3.2 μM). The 7-AdB pre-mRNA carrying inverted repeats (→.←) was also used. (C) Splicing mixtures containing the 7-AdB and the small-intron 7B-Ad were supplemented with His-A1 and GST-A1 (0.5 μM each). (D) The histogram depicts a compilation of three independent experiments performed with the indicated concentrations of recombinant proteins. In each case, individual values obtained for the splicing of 7-AdA (-.-) or 7-AdA (a.a) pre-mRNAs were normalized with the splicing efficiency of the small intron 7B-Ad pre-mRNA. Error bars indicate standard deviations.

previously that an excess of TS10 abrogates the activity of ABS in alternative splicing [6]. An excess of TS10 similarly reduced the splicing efficiency of the 7-AdB(a.a) pre-mRNA, without affecting the amplification of splicing products derived from the 7B-Ad small-intron pre-mRNA (Figure 2A, lanes 6-10). We also tested the impact of increasing the level of hnRNP A1 in the extract by using recombinant A1 protein. GST-A1 stimulated the splicing efficiency of 7-AdB(a.a) (Figure 2B, lanes 5-8). In contrast, splicing of the control small-intron 7B-Ad pre-mRNA was not affected by the addition of GST-A1. GST-A1 also stimulated splicing of the 7-AdB(-.) pre-mRNA in a dose-dependent manner (Figure 2B, lanes 1-4). The reason for this stimulation is unclear. It is possible that weaker ABS in the large intron of the 7-AdB pre-mRNA are activated when the concentration of hnRNP A1 is increased. In contrast, splicing of the large-intron pre-mRNA carrying inverted repeats was not further stimulated by the addition of GST-A1 (lanes 9-12). A His-tagged version of A1 (His-A1) was as active as GST-A1 at stimulating large-intron splicing (Figure 2C and 2D). In contrast, the addition of His-tagged UP1 (His-UP1), a shortened version of A1 lacking the C-terminal glycine-rich domain did not stimulate splicing of the large-intron pre-mRNAs (Figure 2D). Rather, His-UP1 slightly impaired splicing of the ABS-containing 7-AdA(a.a) pre-mRNA, possibly because it antagonized the binding of endogenous hnRNP A/B proteins.

We have recently used antisense oligonucleotides carrying a non-hybridizing ABS-containing tail to alter splicing decisions in vitro and in vivo [27]. The ABS tail interfered with splicing when the antisense portion of the oligonucleotide was complementary to exonic sequences upstream of 5' splice sites. To verify whether ABS-tailed oligonucleotides could now act positively by reproducing the effect of cis-acting

ABS, we added to a HeLa extract a mixture of trans-acting ABS-containing oligos complementary to the ends of the introns in the 7-AdA(-.-) and 7B-AdA(-.-) pre-mRNAs (UA and Da; Figure 3A). The in vitro splicing of these pre-mRNAs was stimulated in both cases (Figure 3B, lanes 1-3 and lanes 9-12, respectively). The UA and Da mixture also stimulated splicing of the 7-AdA(a.a) pre-mRNA (Figure 3B, lanes 4-6). In contrast, the oligo mixture did not improve splicing of a pre-mRNA carrying the B insert (lanes 7-8). In general, concentrations of oligos varying between 0.08 to 160 nM stimulated splicing (representing a molar excess of 10 to 2000-fold relative to the pre-mRNA). The level of stimulation varied between 2 to 8-fold between different experiments (data not shown). Concentrations superior to 160 nM usually promoted a specific reduction in the splicing efficiency of enlarged introns, possibly because of titration of hnRNP A/B proteins by an excess of ABS-containing oligos (data not shown). Our results with the ABS-containing oligo mixture therefore suggest that hnRNP A/B proteins can be recruited at the intended positions in the intron. This conclusion was supported by the results of an immunoprecipitation assay performed in nuclear extracts using an anti-A1 antibody and a portion of the 7-AdA pre-mRNA. The results show that recovery of the RNA is improved by the presence of a cis-acting ABS (Figure 3C, compare lane 3 and lane 4). Likewise, providing a trans-acting ABS as part of the tail of an antisense oligo stimulates recovery (lane 6), a result not observed when the oligo carries a non-ABS tail (lane 5). Thus, a cis- or trans-acting ABS improves the association of hnRNP A1 with the target RNA.

We have proposed that the mechanism underlying the activity of hnRNP A/B proteins in alternative splicing involves an interaction between bound A/B proteins such

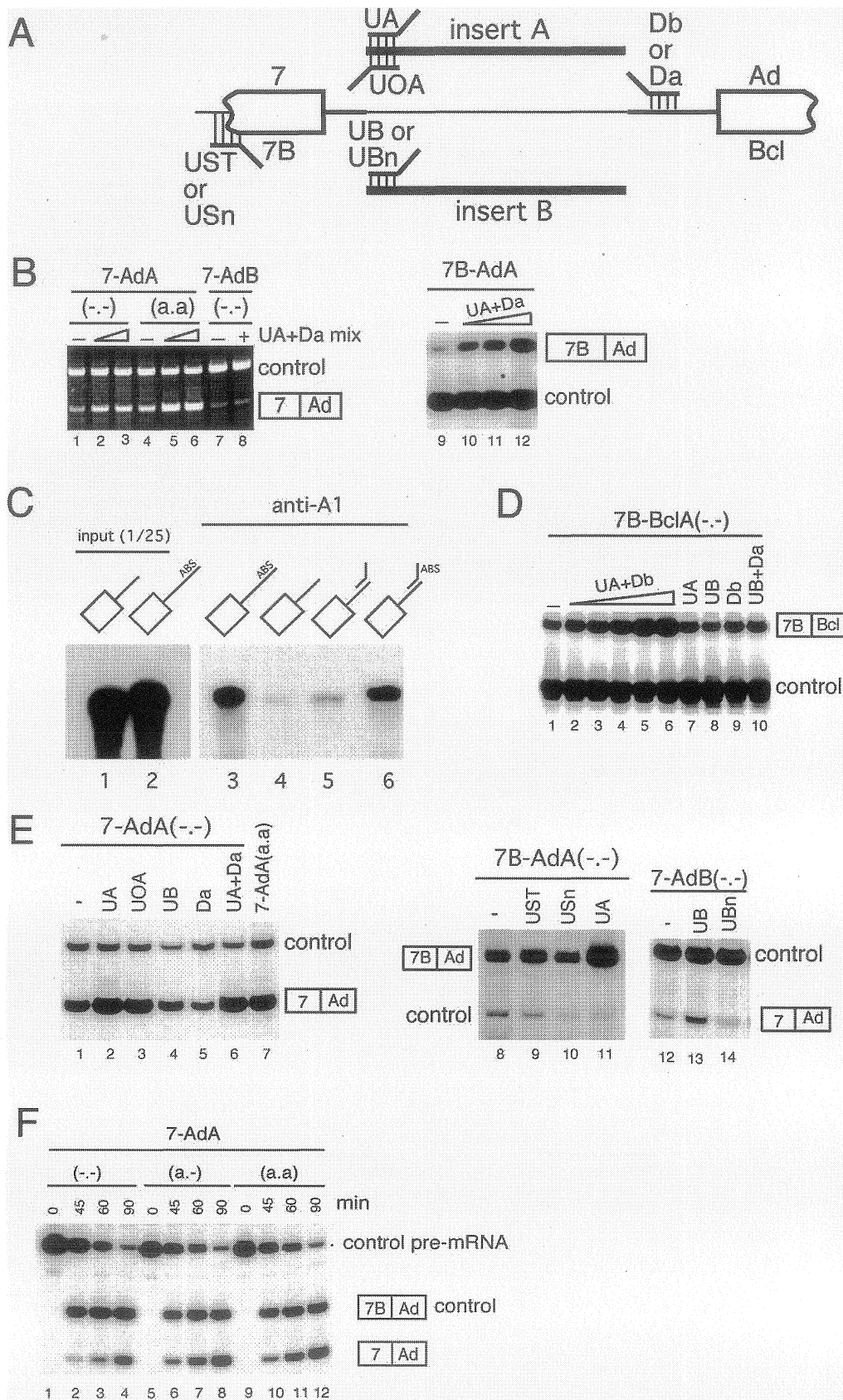


Figure 3

Figure 3. Antisense oligos carrying ABS stimulate the splicing of large introns.(A) Schematic representation of model large-intron pre-mRNAs and the position and structure of the RNA oligos. Oligos UA and UB respectively form a duplex with sequences located 46 to 65 and 46 to 64 nt downstream from the 5' splice site. Oligos Db and Da respectively hybridize 123 to 142 nt and 68 to 87 nt upstream of the 3' splice junction. (B) The 7-AdA pre-mRNA lacking ABS (-.-) or containing ABS (a.a) was incubated in a HeLa extract in the absence (lanes 1 and 4) or in the presence (lanes 2-3 and 5-6) of UA and Da oligos (16 and 40 nM of each oligo). The 7-AdB pre-mRNA (-.-) and the 7B-AdA pre-mRNAs were also incubated in the presence of UA and Da (40 nM in lane 8, and 0.08, 0.8 and 8 nM in lanes 10-12, respectively). As internal control for splicing, a smaller quantity of the small-intron 7B-Ad pre-mRNA (lanes 1-8) or the small-intron 7-Ad pre-mRNA (lanes 9-12) was co-incubated with the test substrates. Incubation in HeLa extracts was for 60 min. (C) Labeled transcripts corresponding to the first 196 nt of the 7-AdA(-.-) pre-mRNA (lanes 1, 4, 5 and 6) or the first 237 nt of the ABS-containing 7-AdA(a.a) pre-mRNA (lanes 2 and 3) were incubated in a HeLa nuclear extract in the presence of 100 pM of the ABS-lacking UAn oligo (lane 5) or the ABS-containing UA (lane 6). Mixtures were immunoprecipitated with an anti-hnRNP A1 antibody and resolved in a denaturing 5% polyacrylamide gel. The initial input for each transcript representing 1/25th of the total amount is shown in lanes 1 and 2. (D) The 7B-BclA was co-incubated with 100-fold less of the small-intron 7-Ad control pre-mRNA and increasing amounts of the UA and Db oligo mixture (0, 0.08, 0.8, 8, 80 and 160 nM) or with 160 nM of individual or mixtures of various oligos. (E) The 7-AdA pre-mRNA was co-incubated with the small-intron 7B-Ad pre-mRNA in a HeLa nuclear extract for 90 min at 30°C. Each oligo was used at a concentration of 160 nM. The 7-AdA(a.a) pre-mRNA containing cis-acting ABS elements was used as a control (lane 7). In lanes 8-11, the 7B-AdA pre-mRNA was co-incubated with the small-intron 7-Ad pre-mRNA and 80 nM of oligos. In lanes 12-14, the 7-AdB pre-mRNA was co-incubated with the 7B-Ad control pre-mRNA and either the UB or UBn (40 nM each). UBn carries a non-ABS tail. Incubation was for 60 min in a HeLa extract. (F) The control 7B-Ad pre-mRNA was co-incubated with 7-AdA pre-mRNA containing either no ABS (-.-), only the upstream ABS (a.-), or two ABS (a.a). Incubation in HeLa extracts was for 0, 45, 60 and 90 min. RT-PCR assays performed with a single pair of primer allows to amplify the unspliced control pre-mRNA, as well as mRNA products derived from both the control and the 7-AdA derivatives.

that portions of the pre-mRNA are looped out, therefore changing pre-mRNA structure to favor contacts between the external pair of exons [6]. One prediction from this model is that the presence of ABS at both ends of the enlarged intron should be required to observe stimulation of splicing. We used trans-acting ABS-containing oligos to assess whether

providing an ABS at either the upstream or the downstream position could stimulate splicing. First, we present the activity of individual trans-acting ABS on a pre-mRNA carrying the 3' splice site of human Bcl-x exon 3 (7B-BclA). The addition of oligos UA and Db stimulated 7B/Bcl splicing (Figure 3D, lanes 1-6), a result that reproduced the activity of cis-acting ABS (Figure 1E). Stimulation required the presence of both oligos since UA or Db alone did not stimulate splicing (Figure 3D, lanes 7 and 9). Providing a non-hybridizing oligo with an ABS tail was inactive (lane 8), and combining this oligo with the non-hybridizing Da oligo also did not provide stimulation (lane 10). Thus, stimulation of 7B-BclA pre-mRNA splicing required a combination of upstream and downstream trans-acting ABS, suggesting cooperative interactions between these sites, in accord with the looping out model. Second, we tested the impact of individual oligos on pre-mRNAs carrying the adenovirus 3' splice site (7-AdA). Notably, the upstream UA oligo alone stimulated 7-AdA splicing nearly as efficiently as the UA and Da mixture (Figure 3E, lanes 2 and 6). The UA oligo alone also stimulated 7B-AdA splicing (lane 11). In contrast, the downstream Da oligo alone did not stimulate splicing (lane 5). Providing the ABS as a 3' rather than a 5' extension was also stimulatory (UOA; Figure 3D, lane 3), and no activity was provided by an oligo carrying a non-ABS extension (UBn; lane 14), or an ABS-containing oligo complementary to the first exon (UST; lane 9). A single cis-acting ABS at the upstream position in the large intron of the 7-AdA also stimulated splicing (7-AdA(a.-); Figure 3F). Trans-acting ABS hybridizing at a distance greater than 250 nt from the 5' splice site did not significantly enhance splicing (data not shown). Likewise, placing two cis-acting ABS in the middle of an enlarged intron did not stimulate splicing (data not shown). Thus, while positioning an ABS relatively close to the 5' splice site is apparently important for splicing stimulation, a single cis- or trans-

acting ABS positioned near the 5' splice site is sufficient for stimulating splicing of the 7-AdA and 7B-AdA pre-mRNAs. This conclusion contrasts with the results obtained with the 7B-BclA pre-mRNA and does not a priori support the looping out model. Although the reason for this difference remains unclear, hnRNP A1 has been reported to bind to the adenovirus 3' splice site [28]. Thus, a high-affinity A1 binding site located at this position may collaborate with an ABS near the 5' splice site to stimulate pre-mRNA splicing. Unfortunately, we could not confirm this hypothesis experimentally because the putative ABS directly overlaps the adenovirus 3' splice site, and mutating the ABS would inhibit splicing.

Antisense oligonucleotides carrying ABS stimulate splicing of an enlarged intron in vivo

To address whether ABS can stimulate intron splicing in vivo, we relied on a model pre-mRNA used previously to demonstrate the negative influence of intron size on exon inclusion [2]. The CD44 model pre-mRNA contains the V3 alternative exon flanked downstream by an enlarged intron containing six adjacent 977 nt-long spacer elements derived from lambda DNA (Figure 4A). In this configuration, the enlarged intron promotes exon V3 skipping [2]. We tested the effect of providing trans-acting ABS at one or both ends of the enlarged intron. Following transfection of the CD44 plasmid in COS-7 cells, a second transfection was performed with antisense 2'O-Me oligonucleotides carrying an ABS tail. hnRNP A1 binds very well to 2'O-Me RNA [27]. Compared to a control oligo lacking an ABS tail (UV3NT; Figure 4B), oligos complementary to the 5' end of the intron and carrying only one ABS stimulated exon V3

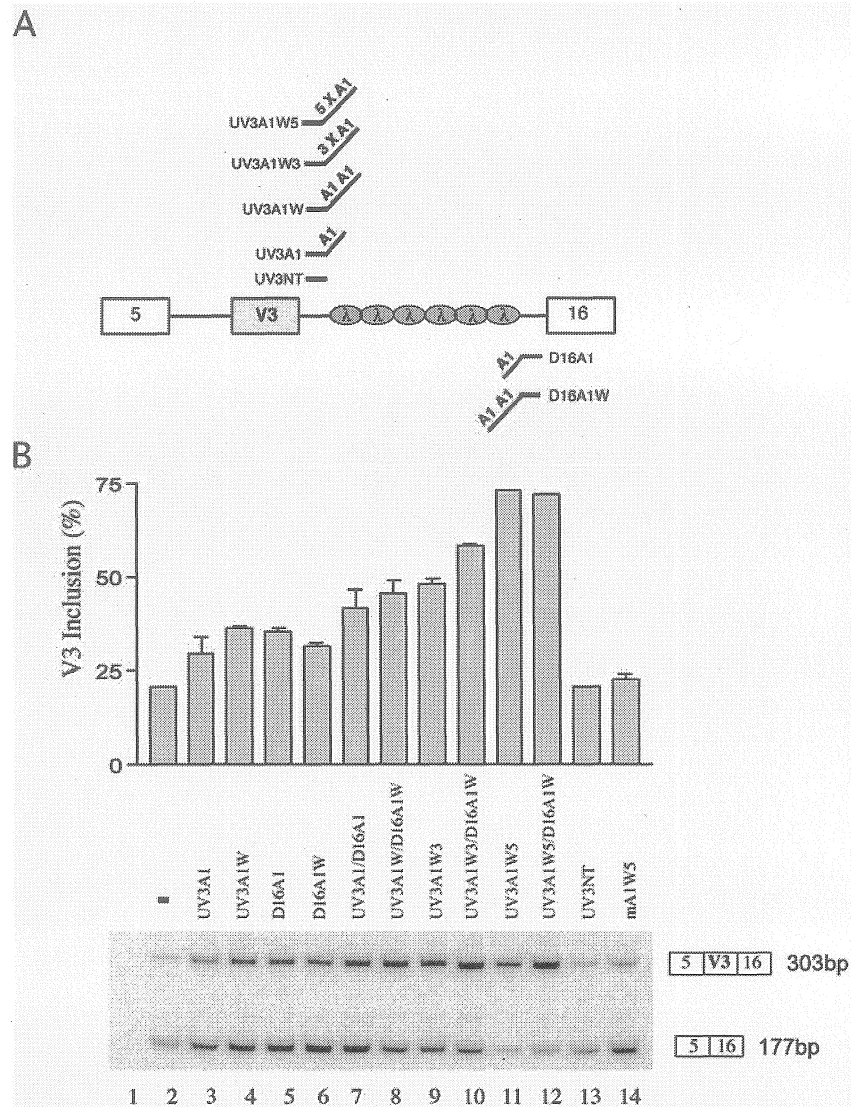


Figure 4. The in vivo splicing of a large intron is improved by ABS-containing oligos. (A) Structure of the p44:v3λλλλλλ pre-mRNA. This pre-mRNA contains the alternatively spliced V3 exon as well as constitutive exons 5 and 16 from the CD44 gene. Six 1-kb lambda DNA repeats (gray circles) were inserted downstream of V3 to increase the length of this intron [2]. The position and structure of the RNA oligos are depicted. A1 indicates an ABS element. (B) COS-7 cells were transfected with plasmid p44:V3λλλλλλ. Twenty-four hours later they were treated with different 2'O-Me RNA oligos, and total RNA was extracted after 24 h. A RT-PCR assay was performed in the presence of [32 P]dCTP to determine the relative levels of both mRNA splicing products. The inclusion frequency of exon V3 expressed as a percentage is shown graphically with mean value and error bars derived from three separate experiments.

inclusion very slightly (from 20% with UV3NT to 29 % with UV3A1 with a p value of 0.1 based on three separate experiments). A more important stimulation was obtained when the tailed oligonucleotide carried two, three or five ABS (UV3A1W; UV3A1W3 and UV3A1W5, respectively promoting an average of 36%, 48% and 73% inclusion). A control oligo with five adjacent ABS but lacking a portion complementary to the CD44 mini-gene did not improve exon V3 inclusion (mA1W5). Exon inclusion was also stimulated by tailed oligonucleotides complementary to the 3' portion of the intron (D16A1 and D16A1W). Providing oligos as mixtures (UV3A1/D16A1, UV3A1W/D16A1W, UV3A1W3/D16A1W or UV3A1W5/D16A1W) offered little or no additional stimulation when compared to the effect of providing the upstream oligo alone. Consistent with our previous study [27], the activity of the ABS-tailed oligos was compromised when the concentration of hnRNP A1/A2 proteins was reduced through the use of siRNAs targeting the A1/A2 mRNAs (data not shown). Thus, oligonucleotides designed to deliver an ABS at the 5' or the 3' end of the intron stimulated splicing of the enlarged intron in vivo. However, this experiment did not reveal an apparent cooperation between terminal ABS (see Discussion).

hnRNP F/H binding sites duplicate the in vitro modulating activity of hnRNP A/B binding sites

hnRNP F/H proteins have affinity for G-stretches [23]. Thus, a subset of the GGG motifs found near the ends of mammalian introns may be bound by hnRNP F/H proteins. Moreover, hnRNP F/H proteins contain glycine-rich domains, and similar domains promote an interaction between hnRNP A1 proteins. For these reasons, we tested whether a binding site for hnRNP F/H proteins (FBS) could stimulate the in vitro splicing of an

enlarged intron. We used the hnRNP H binding site identified in the cystathionine β -synthase gene [29], a site that contains 2 G quadruples (Figure 5A). The insertion of two copies of this FBS at the upstream position in the intron of the 7-AdB pre-mRNA did not stimulate in vitro splicing (Figure 5B, lanes 2 and 4). Likewise, inserting one FBS at the downstream position of the enlarged intron did not improve splicing efficiency (Figure 5B, lane 6). However, when both the upstream and the downstream FBS were present, splicing was strongly stimulated (lane 8).

To further explore the ability of FBS to mimic ABS, FBS were inserted into a model pre-mRNA to monitor the impact of FBS on 5' splice site selection. We used the 553 pre-mRNA which contains the 5' splice sites of hnRNP A1 exons 7 and 7B joined to the downstream adenovirus 3' splice site [7]. Inserting one FBS immediately downstream of exon 7 or immediately downstream of exon 7B only slightly stimulated distal 5' splice site selection (553f- and 553-f; Figure 5C, lanes 1-3). In contrast, the presence of FBS at both positions promoted a strong increase in distal 5' splice site usage (553ff; Fig 5C, lane 4), thereby duplicating the impact of ABS in 5' splice site selection [7]. Thus, the FBS elements functioned cooperatively in vitro to stimulate distal 5' splice site selection and splicing of an enlarged intron.

The contribution of hnRNP F/H proteins in the activity of the FBS element was confirmed in various ways. First, a gel-shift assay indicated that recombinant hnRNP H protein but not hnRNP A1 bound the FBS element (Figure 6A, lanes 1-7). In contrast, an oligo carrying two copies of the A1 binding motif used in ABS (UAGAGU) had more affinity for hnRNP A1 than hnRNP H (lanes 8-14). Second, the distal 5' splice site promoting activity of FBS was dependent upon hnRNP F/H proteins. This was shown by

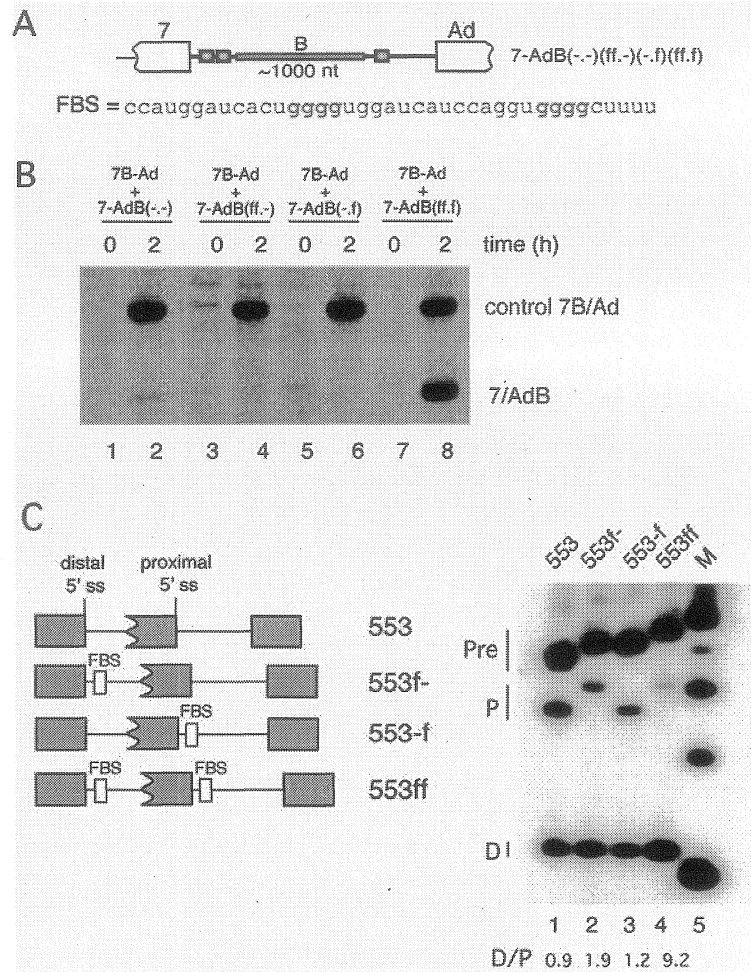


Figure 5. Binding sites for hnRNP F/H proteins stimulate the splicing of an enlarged intron and modulate 5' splice site selection. (A) Structure of the pre-mRNA containing the enlarged intron (lambda insert B). The hnRNP F/H binding site (FBS) was described previously [29]. The first 6 nt of the FBS are derived from a NcoI site used for cloning purposes. (B) The control 7B-Ad pre-mRNA (8.7 pM) and the 7-AdB pre-mRNAs (85 pM) containing either no FBS (-.-), two FBS at the upstream position (ff.-), one FBS at the downstream position (-.f), or FBS at both positions (ff.f) were co-incubated for the times indicated (in hours) in a HeLa nuclear extract. A RT-PCR assay was performed to amplify simultaneously spliced mRNAs derived from 7B-Ad or the 7-AdB derivatives. (C) Activity of FBS in 5' splice site selection. The structure of the pre-mRNA containing a FBS at one or both positions is shown on the right. The pre-mRNAs were incubated in a HeLa extracts for 2 hours. A RT-PCR assay was performed to amplify splicing products from splicing mixtures. M = molecular weight markers. The position of the products generated from the use of the distal (D) and proximal (P) 5' splice sites as well as from the pre-mRNAs (Pre) is shown. The ratio of the distal/proximal products (D/P) is indicated below the lane number.

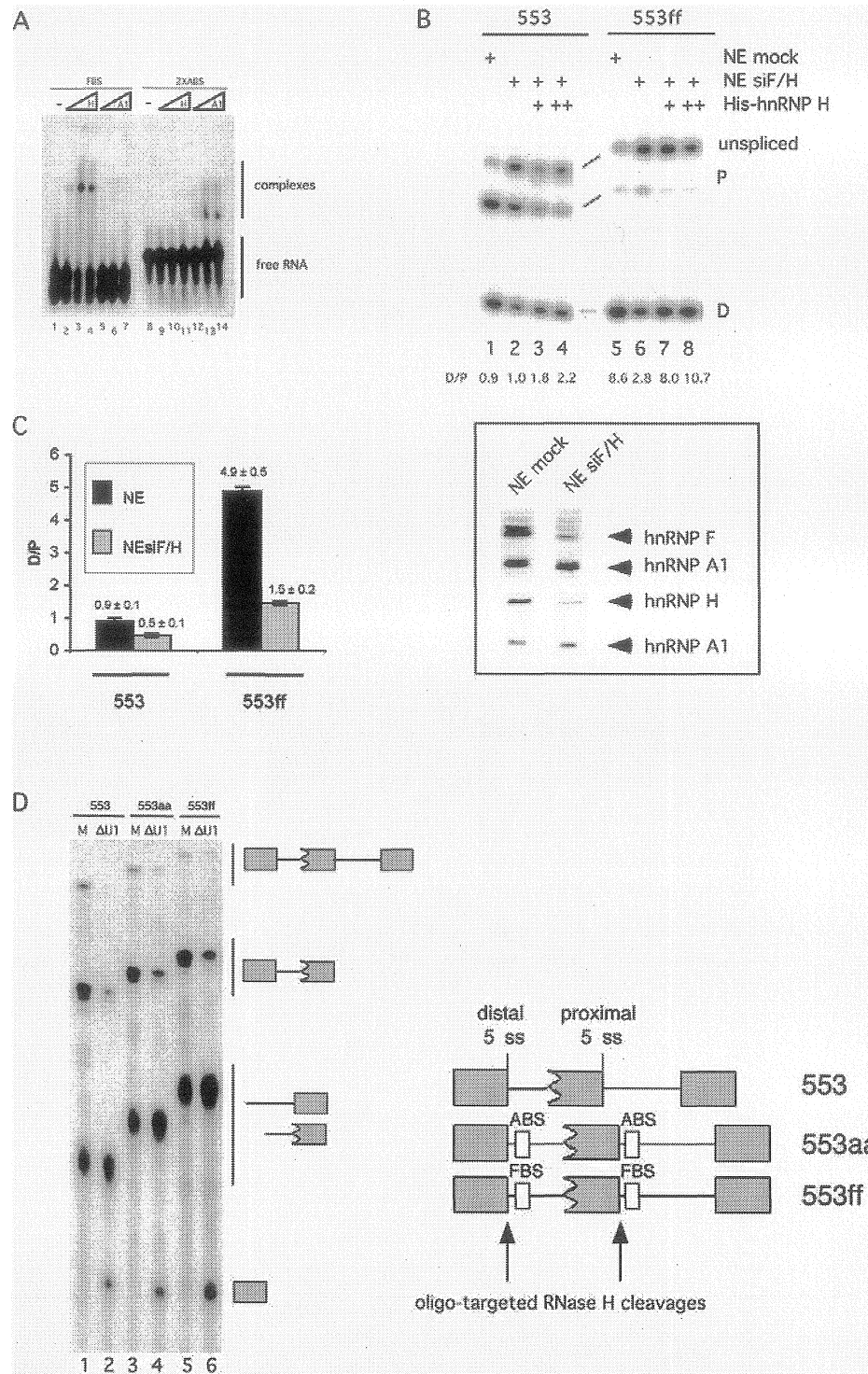


Figure 6. Role of hnRNP F/H proteins in the activity of FBS. (A) hnRNP H binds FBS specifically. The binding of hnRNP H and hnRNP A1 to FBS and ABS was monitored by using a gel-shift assay. The sequence of the FBS RNA corresponds to the sequence shown in figure 4A, while 2XABS is a 2'-O-Me oligo that contains two UAGAGU

elements (CCUUUAGAGUAGUAGAGUAGAAUAAG-CCUUGCAUAAAUGG). Binding conditions were as described [43] and used 1.25, 2.5 and 3 μ M of hnRNP H or A1 proteins. (B) hnRNP F/H are required for the activity of FBS on 5' splice site selection. Nuclear extracts were prepared from HeLa cells that were treated with siRNAs against hnRNP F/H [34]. Pre-mRNA substrates lacking or containing FBS were assayed in extracts prepared from mock-treated and siF/H-treated cells. The siF/H extract was also supplemented with recombinant His-tagged hnRNP H protein prepared from baculovirus-infected cells (0.15 μ M). The ratio of the products resulting from the use of the distal or proximal 5' splice site is indicated below the lane number. The right panel shows a western analysis of the content of hnRNP F and H proteins in extracts prepared from mock-treated and siF/H-treated cells. In addition to the anti-F or anti-H antibody, an anti-A1 antibody [45] was co-incubated to reveal A1 and monitor total protein loading. (C) Splicing assays using the 553 and the 553ff pre-mRNAs were performed in triplicates in extracts prepared from mock-treated and siF/H-treated HeLa cells. The ratio of the amplified products corresponding to the proximal and distal 5' splice site usage was calculated and plotted in a graph that displays error bars. (D) Oligonucleotide-mediated RNase H protection assays to monitor U1 snRNP occupancy on the competing 5' splice sites. Pre-mRNAs lacking or containing FBS or ABS were incubated at 0°C in mock-treated and U1 snRNP-inactivated extracts (Δ U1). Oligonucleotides complementary to the 5' splice sites were added along with RNase H. The position of the fully protected pre-mRNAs and cleaved molecules is shown.

incubating the 553ff pre-mRNA in a nuclear extract prepared from HeLa cells that had been treated with siRNAs to knockdown hnRNP F/H expression (Figure 6B, right panel). The distal/proximal ratio of products derived from the 553ff pre-mRNA was decreased in such an extract, whereas this ratio was minimally affected with the 553 pre-mRNA (Figure 6B, lanes 5-6 and lanes 1-2, respectively; see also Figure 6C). Moreover, while the addition of recombinant hnRNP H protein only had a modest effect on 553 pre-mRNA splicing, the distal/proximal ratio of products derived from 553ff pre-mRNA splicing was improved when hnRNP H was added to the siF/H extract (Figure 6B, lanes 3-4 and 7-8, respectively). These results suggest that the activity of the FBS in the HeLa extract requires at least the hnRNP H protein. We further asked whether the strong shift in 5' splice site selection mediated by the pair of FBS was associated with a corresponding

change in the binding of U1 snRNP to 5' splice sites. To assess U1 snRNP binding we performed an oligonucleotide-mediated RNase H protection assay using a mixture of oligonucleotides complementary to the 5' splice sites of exon 7 and exon 7B [24]. The protection profile obtained at 0°C with the control 553 pre-mRNA indicates two populations of U1 bound-pre-mRNAs; one bound only to the distal 5' splice site, and a less abundant population (15%) to which U1 is bound to both the distal and the proximal 5' splice sites (Figure 6D, lane 1). For pre-mRNAs carrying ABS or FBS, the percentage of transcripts bound by U1 at both the distal and the proximal 5' splice sites was reduced slightly (11% and 4%, respectively; Figure 6D, lanes 3 and 5). The protection observed in all cases was largely U1 snRNP-dependent (lanes 2, 4 and 6). Our results therefore indicate that the robust improvement in distal 5' splice site usage mediated by FBS and ABS was not accompanied by equivalent changes in U1 snRNP binding to the competing 5' splice sites.

DISCUSSION

Stimulation of splicing by intronic binding sites for hnRNP A/B and hnRNP F/H proteins

The initial demonstration that hnRNP A1 antagonized the activity of SR proteins in splice site selection assays was followed by many reports implicating the hnRNP A/B proteins in the activity of exonic silencer elements. Understandably, these findings led hnRNP A/B proteins to be regarded mostly as negative regulators of splicing. The results presented here suggest that the binding of hnRNP A/B proteins in introns can also play a positive role in the generic splicing reaction. Using model pre-mRNAs harboring artificially enlarged introns that are spliced poorly in HeLa nuclear extracts, we have shown that intronic high-affinity A/B binding sites (ABS) positioned near splice junctions can stimulate *in vitro* splicing.

On the other hand, hnRNP F/H proteins have been implicated in the activity of both splicing enhancers and silencers. For example, the hnRNP F and H proteins are part of a complex assembling on an intronic enhancer element that promotes the neuro-specific inclusion of the N1 exon in the *src* pre-mRNA [21,30]. hnRNP H also activates an SC35-bound exonic enhancer element in the human immunodeficiency virus [31], but is required for the activity of a silencer element located in a rat β -tropomyosin alternative exon [32]. Binding sites for hnRNP H that overlap 5' or 3' splice sites can also repress splicing [16,29,33]. Recently, we have uncovered a positive role for hnRNP F/H proteins when bound downstream of the Bcl-x_s 5' splice site [34]. We now add to this list of activities the observation that binding sites for hnRNP F/H located at the ends of an

enlarged intron can stimulate *in vitro* splicing. The ability of FBS to replicate the activity of ABS is not limited to the splicing of enlarged intron since FBS also promoted distal 5' splice site utilization. This situation contrasts with a recent report documenting a complex interplay between exonic hnRNP A1 binding sites and an intronic GGGG motif in the inclusion of the brain-specific GRIN1 CI exon [16]. In this case, hnRNP H binding to a GGGG motif appears to antagonize the silencing activity of exonic A1 binding sites. These results suggest that hnRNP A1 and H may exhibit different roles depending on the precise arrangement of their respective binding sites relative to a 5' splice site.

hnRNP proteins and the looping out model

Importantly, upstream and downstream binding sites for hnRNP F/H or hnRNP A/B proteins displayed an apparent cooperativity in some pre-mRNAs. A pair of ABS was required to stimulate splicing of the 7B-BclA pre-mRNA. A similar requirement was noted for FBS to stimulate the splicing of an enlarged intron. These results are consistent with the view that the mechanism underlying the stimulatory activity of hnRNP A/B and F/H proteins involves an interaction involving terminally bound proteins such that a portion of the intron is looped out to bring in closer proximity distantly separated exons (Figure 7). The situation was different when we used pre-mRNAs carrying the adenovirus major late 3' splice site. In this case, an ABS positioned near the 5' splice site was sufficient for stimulation. Although we cannot rule out alternative explanations, the fact that the adenovirus 3' splice site is bound by hnRNP A1 [28] may explain why an upstream ABS is sufficient for stimulating splicing of pre-mRNAs carrying this 3' splice site. Cooperativity was also not observed when we targeted the enlarged intron of flanking CD44 exon V3 *in vivo*. In this case, we noted that each of the six 977-long

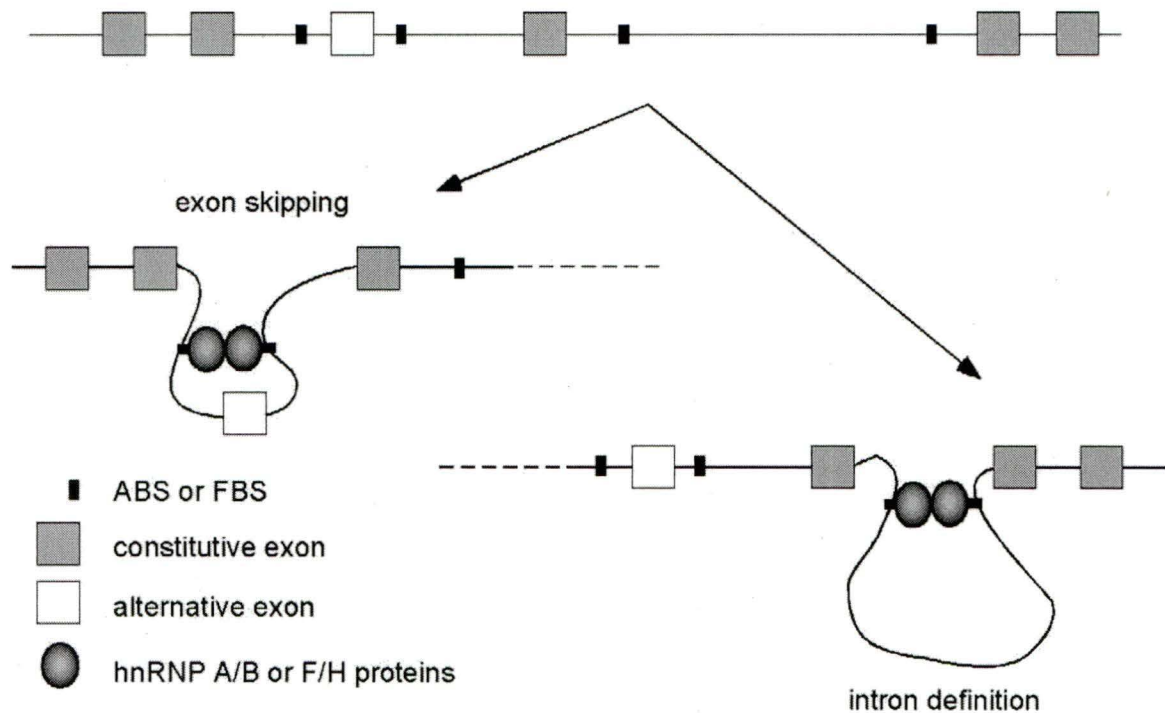


Figure 7. The looping-out model of action for hnRNP A/B and hnRNP F/H proteins. hnRNP proteins bound to high-affinity binding sites (ABS or FBS) would self-interact to loop out intron sequences and stimulate intron definition. A similar interaction involving ABS or FBS located in distinct introns would loop out an alternative splice site or a cassette exon to favor skipping and commitment between the external pair of splice sites. It remains unknown whether heterotypic interactions can occur between hnRNP A/B and hnRNP F/H proteins.

lambda insert contain a putative ABS (UAGGGU) at position 666. Thus, we propose that an internal ABS may cooperate with a terminal trans-acting ABS to stimulate intron removal.

The activity of hnRNP A/B and hnRNP F/H binding sites that is documented here is relevant to the reported activity of GGG motifs which are found more abundantly near the ends of introns [15]. Such motifs have been associated with an ability to modulate 5' splice site selection and facilitate intron splicing [13,15,17,18,19]. All binding sites for

hnRNP F/H proteins characterized so far contain GGG, but it remains unclear whether all GGG motifs are bound by hnRNP F/H proteins since U1 snRNP and SF1 have also been proposed to bind to GGG motifs [18,35]. We have tested hnRNP A1, F and H binding to the two G-rich elements that improve splicing of a *Fugu* intron in mammalian cells [15]. As judged by gel-shift analysis, the two GGG motifs in G1 are not bound by F, H or A1, whereas the G2 element, which contains a G triple and a GGGG, is bound strongly by hnRNP H but not hnRNP F nor hnRNP A1 (Figure S2). Although the winner high-affinity site for A1 contains GGG [20], the A1 binding site that we have used (UAGAGU) lacks a G triplet and is not bound by hnRNP H. Thus, many GGG motifs found near the ends of introns may be binding sites for hnRNP A/B and F/H proteins, and they may contribute to intron definition.

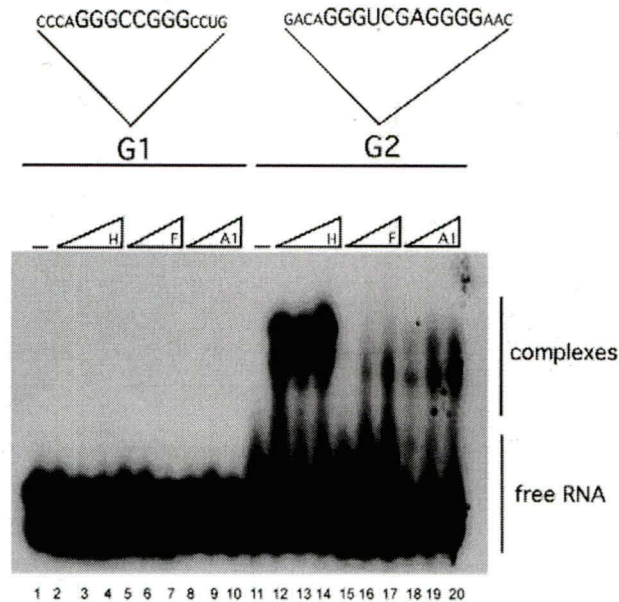


Figure S2. Gel-shift assay using recombinant hnRNP H, F and A1. The initial G1 and G2 elements (larger letters) stimulated splicing when inserted into a *Fugu* intron. The sequences immediately flanking the insertion sites are also shown. The 16 nt G1 and 19 nt G2 RNA oligonucleotides were 5' end labeled and individually incubated with hnRNP proteins (1, 2 and 3 μM) in the presence of heparine. Complexes were fractionated on a native 5% acrylamide gel.

According to the simplest version of the looping out model, hnRNP A/B or F/H proteins bound near the ends of introns would interact with one another to loop out most of the intron. It is possible however that proteins bound to terminal ABS or FBS also interact with proteins bound to internal sites to loop out portions of the intron. In this case, terminal ABS may insure a maximal effect on intron definition. To examine the generality of the potential importance of terminal ABS and FBS, we performed a computational analysis for the presence of intronic ABS and FBS at the ends of 156525 human introns of sizes ≥ 330 nt (Table I). Consistent with previous studies, we find that introns carrying one or more GGG motifs near the 5' splice site (from positions +11 to +150, relative to the 5' splice junction) or near the 3' splice site (from positions -41 to -180, relative to the 3' splice junction) are significantly overrepresented when compared to the values obtained if the sequence in these regions is randomly shuffled (Table I, sh).

		Intronic motifs near splice sites			
		GGG	GGGG (FBS)	AGGG ^U / _A (ABS)	ACAC
Total 5'ss		116207 (74%)	64382 (41%)	48413 (31%)	45683 (29%)
	sh	104675 \pm 219 (67%)	50721 \pm 178 (32%)	32115 \pm 171 (21%)	44046 \pm 201 (28%)
Total 3'ss		111467 (71%)	54268 (35%)	45343 (29%)	45820 (29%)
	sh	101032 \pm 209 (65%)	43842 \pm 176 (28%)	31897 \pm 169 (20%)	46333 \pm 185 (29%)
Neither		17382 (11%)	68376 (44%)	79292 (51%)	79096 (51%)
	sh	23914 \pm 128 (15%)	81799 \pm 149 (52%)	99889 \pm 182 (64%)	79614 \pm 198 (51%)
Both ends		88531 (57%)	30501 (19%)	16523 (11%)	14074 (9%)
	sh	73096 \pm 161 (47%)	19837 \pm 117 (13%)	7376 \pm 82 (5%)	13468 \pm 120 (9%)
	pr	82606 (53%)	22269 (14%)	14005 (9%)	13337 (9%)

Table I. Overrepresentation of human introns carrying FBS or ABS at both ends.

The presence of GGG, GGGG (FBS), AGGG^U/_A (ABS) and ACAC motifs in 158987 human introns was compiled from positions +11 to +150 (relative to the 5' splice site) and -41 to -180 (relative to the 3' splice site). The number of introns (and relative percentage) containing at least one motif at the upstream or the downstream position (5'ss and 3'ss, respectively), no motif at either position (Neither) or motifs at the two positions (Both ends) is indicated in black. Shuffled portions were used to calculate how many introns (sh) carried motifs at similar positions based on random occurrence (numbers in red represent mean values with standard deviations). The predicted (pr) set given for introns carrying motifs at both ends (in blue) was calculated from the actual fraction of introns carrying motifs at individual ends. If p_5 is the fraction of introns having a 5' motif and p_3 is the fraction of introns having a 3' motif, then the 'predicted' number of introns having both (assuming independence of the occurrences on both ends) would be $p_3 \times p_5$.

Next, we asked whether there was a similar enrichment for introns carrying GGG motifs at both ends of the same intron. Interestingly, 88531 introns carried at least one GGG at both ends, a number significantly above the predicted number (82606) based on the actual number of introns with GGG motifs at either the 5' or the 3' end (Table I, pr). To discriminate between hnRNP F/H binding sites and hnRNP A/B binding sites, we repeated the analysis using the GGGG motif as a FBS, and the AGGG^U/_A motif as an ABS. Notably, 30501 introns carried at least one GGGG motif at both extremities, representing an excess of 8232 introns relative to the number predicted if occurrences at each end were unlinked. A slightly less important enrichment was observed when the analysis was performed with the ABS motif. In this case, 16523 introns contained at least one ABS motif at both ends, representing an excess of more than 2518 introns. As a control, we carried out the analysis with the sequence ACAC. Introns carrying this motif at one or both ends were not significantly overrepresented (Table I). Thus, our results are consistent with the view that ABS and FBS present at both ends of introns cooperate to

function in splicing. No bias in the configuration of these motifs according to intron length was noted (not shown), suggesting that cooperative interactions involving terminal ABS or FBS may occur in a large fraction of introns irrespective of their sizes.

The looping-out model is also the simplest way to explain the behavior of ABS and FBS in 5' splice site selection. A pair of FBS was considerably more active than individual FBS at shifting splicing towards the distal 5' splice site (Figure 5C). Such cooperativity was not observed for ABS since distal 5' splice site usage was significantly stimulated by positioning an ABS downstream of either the proximal or the distal 5' splice site [7]. Because this pre-mRNA contains the same adenovirus 3' splice site as the one used in large intron substrates, a contribution of the ABS at the 3' splice site may also contribute to explain this apparent lack of cooperativity. Thus, appropriately positioned ABS or FBS can promote alternative 5' splice site usage and, possibly, exon skipping (Figure 7). Consistent with this view, we have shown that deleting the intronic ABS flanking constitutive exon 7 and/or alternative exon 7B in the hnRNP A1 pre-mRNA promotes exon 7B inclusion [6].

Additional experimental evidence indirectly support the looping out model. A1 proteins self-interact [36], and A1 molecules bound to one ABS can simultaneously interact with another ABS [6]. As these activities require the glycine-rich domain of hnRNP A1, it is notable that hnRNP F/H proteins also contain glycine-rich domains, and that an interaction between hnRNP F and H has been described [30]. The crystal structure of UP1 bound to high-affinity DNA binding sites is also consistent with the looping model [37]. UP1 contains the two RNA recognition motifs (RRMs) but lacks the C-terminal glycine-rich domain of A1. In the co-crystal, UP1 exists as a dimer, the RRM1 of

each UP1 molecule being bound to sites located on two distinct oligonucleotides. Although UP1 lacks the activity of A1 in splicing and does not self-interact in biochemical and two-hybrid assays [36], the very high concentration of proteins used to promote crystal formation may have forced UP1 dimerization. Thus, we envision that A1 proteins individually bound to distinct high-affinity sites may interact through their glycine-rich domains. This A1/A1 interaction would bring in close proximity the two RNA regions, and the RRM2 domains of each A1 molecule may subsequently cross-interact with sequences flanking the other ABS to stabilize the complex. We speculate that the RRM2 domain of A1 may engage in this type of interaction because the RRM1 is sufficient for specific binding to one high-affinity site [38]. Finally, we have shown that providing FBS or ABS near the ends of an enlarged intron is functionally equivalent to having inverted repeats at these positions. Duplex-forming elements are present in many yeast introns to facilitate commitment between pairs of splice sites [39,40,41,42]. It is intriguing to consider that interactions between hnRNP proteins, rather than duplex formation, may have been selected to help defining mammalian introns. However, it is possible that, in some situations, base-pairing interactions between sequences flanking individual ABS or FBS may further stabilize the conformational changes initiated by hnRNP proteins.

According to the looping-out model, the interaction between hnRNP A/B or between F/H proteins bound in the intron near splice junctions would represent a key step leading to efficient commitment complex formation. In vitro, the splicing efficiency of a small intron was not affected by ABS or by variations in the concentration of A1, possibly because the splice site pairing step is not rate-limiting. Although the presence of ABS and

FBS improved the splicing efficiency of enlarged introns in vitro, the activity of these elements may be more important in vivo when splicing decisions are taken co-transcriptionally. The hnRNP A/B or F/H-mediated looping-out of intron sequences as they exit from the RNA polymerase II transcription complex may facilitate intron definition by improving the frequency of an encounter between a U1-bound 5' splice site and a U2AF-bound 3' splice site located downstream from the ABS/ABS or FBS/FBS complex. The looping out of an intron may also occur in several steps, as would be expected when additional ABS or FBS are distributed along an intron. Given that a 5' splice site located in between two ABS is repressed [7], the ABS- or FBS-mediated looping out of portions of introns may neutralize a multitude of weaker and/or non-productive interactions with pseudo or cryptic splice sites. However, if a splice site located between hnRNP binding sites is strong enough, its commitment to an upstream 5' splice site or a downstream 3' splice site may kinetically outcompete the looping out process mediated by hnRNP proteins. The relative frequency of the two events would be expected to contribute towards setting alternative splicing profiles, and hence, may be influenced by the position of the splice sites relative to ABS or FBS, the speed of transcription and the presence of silencers/enhancers flanking alternative splice sites. Finally, the mechanism by which a looped out splice site is repressed by flanking ABS or FBS remains unclear. The binding of U1 snRNP to a 5' splice site is not greatly affected by flanking ABS or FBS. However, this mechanism of repression may be similar to the mechanism by which inverted repeats repress 5' splice site usage when substituted for ABS or FBS [7]. Commitment, or a later step of spliceosome assembly, may be compromised or delayed by a relatively rigid complex (ABS/ABS, FBS/FBS or a duplex

structure) whose topology may be incompatible with the structural flexibility necessary for efficient spliceosome assembly.

We are currently investigating whether heterotypic interactions can take place between hnRNP A/B and hnRNP F/H proteins as well as other glycine-rich containing RNA binding proteins. If so, such interactions may also play a role in remodeling the conformation of mammalian pre-mRNAs, with a significant impact on splicing efficiency and splice site selection.

MATERIALS AND METHODS

Oligonucleotides

The DNA primers used for the RT-PCR amplification of spliced products were 20 nt-long. E-Ad and BclX3 were used as downstream primers for the RT step and the PCR amplification of products carrying the adenovirus or Bcl-x as second exon, respectively. E-Ad (5'-GAGTTTGTCTCAACCGCGA-3') is complementary to the 5' end of the adenovirus exon L2. BclX3 (5'-TCGGCTGCTGCATTGTTCCC-3') is complementary to a region in Bcl-x exon 3. The upstream primer in all amplifications from in vitro splicing assays was a 21 nt-long oligo T3-5' (5'-GGGAACAAAAGCTGGGTACCG-3') that hybridizes near the 5' end of all transcripts synthesized from the T3 RNA polymerase promoter.

Custom-made RNA oligos were purchased from Dharmacon Research Inc. (Lafayette, CO, USA). The 3' half of the upstream oligo UA or UB is complementary to the intronic sequences at the 5' end of the lambda insert A or B, respectively, 42 nt downstream from the 5' splice site. These oligos have a CE1a element sequence at the 5' end portion. Oligo UOA contains the same complementary sequences but the CE1a element is located at the 3' end. The downstream oligos Da and Db are complementary to a 20 nt region 67 nt upstream of the adenovirus exon L2, and 122 nt upstream of the Bcl-x exon 3, respectively. These oligos contain the CE1a element at their 3' end portion. Oligo UBn shares its last 19 nucleotides with oligo UB but has a non-ABS 25 nt-long tail at its 5' end. Oligo UST has a 20 nt at the 3' end complementary to the intronic sequences

between the distal and the proximal 5' splice sites in RNA 53 [7] while the 5' portion of this oligo contains the CE1a element.

2'O-Me oligos used in transfection assays include UV3A1, which is complementary to a 20 nt region starting 20 nt downstream from the 5' splice site of exon V3. UV3A1 has the ABS from the CE1a element [43]. UV3NT contains the same complementary sequences as UV3A1 but lack the ABS. UV3A1W, UV3A1W3 and UV3A1W5 respectively contain two, three and five consecutive ABS derived from the winner binding site for hnRNP A1 [20]. D16A1 and D16A1W respectively contain the CE1a and the winner A1 binding sites. mA1W5 carries five ABS but is complementary to the pre-mRNA of hnRNP A1. The sequences of all oligos used in splicing are shown in Table II.

Transcription and splicing assays

Constructs containing the adenovirus exon L2 were linearized with ScaI whereas constructs containing Bcl-x exon 3 were linearized using BglII, and used as templates for in vitro transcription. In general, pre-mRNA substrates were synthesized in vitro using T3 RNA polymerase (USB) in the presence of minimal amounts of ³²P-UTP and gel-purified as described [7]. A known amount of the pre-mRNA was then incubated in HeLa nuclear extract [44] under standard splicing conditions [24] at 30°C. The RNA material was then extracted with phenol-chloroform-isoamylalcohol (PCA) and ethanol precipitated. To investigate the effect of RNA oligos on splicing, pre-mRNA molecules were mixed with

Oligo	Length (nt)	Sequence (5'-3')
UA	40	GGGUACCUUU UAGAGU AGGCC <u>CGCUGCGUGAGUAUCCGUGA</u>
UB	40	GGGUACCUUU UAGAGU AGGCC <u>UCGGCUUGGUGUUCUUUCAG</u>
UOA	40	<u>CGCUGCGUGAGUAUCCGUGAG</u> GGGUACCUUU UAGAGU AGGCC
UBn	45	guucgaucucguaacgaaggcguaCGGCUUGGUGUUCUUUCAG
Da	40	<u>GACGUGCAGGUCAAGCUUGA</u> GGGUACCUUU UAGAGU AGGCC
Db	40	<u>CUCUGGGCCAGGUAAAAGGGC</u> GGGUACCUUU UAGAGU AGGCC
UST	40	GGGUACCUUU UAGAGU AGGCCUCCUGUCCACCAGGGCUGCA
USn	45	guucgaucucguaacgaaggcguaGCUGUCCACCAGGGCUGCACC
UAn	40	guucgaucucgauacgaucgCGCUGCGUGAGUAUCCGUGA
UV3A1	40	GGGUACCUUU UAGAGU AGGCCAAUAAGCCUUGCAUAA <u>AUGG</u>
D16A1	40	<u>CAGUUACAUCUGCUGUGAUG</u> GGGUACCUUU UAGAGU AGGCC
UV3NT	20	<u>AAUAAGCCUUGCAUAA<u>AUGG</u></u>
UV3A1W	40	UAUGAU AGGGACU U AGGGUG AAUAAGCCUUGCAUAA
D16A1W	40	<u>CAGUUACAUCUGCUGUGAUG</u> UAUGAU AGGGACU U AGGGUG
UV3A1W3	45	GGU AGGGUCU U AGGGUCU U AGGGUC AAUAAGCCUUGCAUAA <u>AUGG</u>
UV3A1W5	60	GU AGGGUCU U AGGGUCU U AGGGUCU U AGGGUCU U AGGGUC AUAAGCCUUGCAUAA <u>AUGG</u>
mA1W5	60	GU AGGGUCU U AGGGUCU U AGGGUCU U AGGGUCU U AGGGUC ACCUCGCGUGAGUAUCC

Table II. Sequence of the antisense 2'-O-Me RNA oligonucleotides. The antisense portions are underlined and the A1/A2 binding sites are in bold. The non-ABS extensions of UBn, USn and UAn are shown in small case letters.

either the individual oligo or a mixture of the oligos prior to splicing. RNA species obtained after splicing were quantitated and resuspended in sterile water to a

concentration of 5-10 attomoles per μl . An equivalent amount of this solution was then subjected to RT-PCR amplification. To analyze pre-mRNA splicing on conventional denaturing acrylamide gels, uniformly labeled pre-mRNAs were synthesized and processed as described [7].

The treatment of HeLa cells with siRNA to knockdown hnRNP F/H was performed as described [34]. Western analysis was performed using antibodies against hnRNP F and hnRNP H (kindly provided by Douglas Black). Recombinant his-tagged hnRNP F and hnRNP H proteins were produced from baculovirus-infected cells as described [34].

RT-PCR

The pre-mRNAs incubated in splicing extracts were minimally labeled such that the amount of pre-mRNA used could be precisely quantitated and followed until after PCA extraction and ethanol precipitation. In many experiments, a small-intron pre-mRNA was co-incubated with the test pre-mRNA in splicing mixtures to assess equivalent recovery and loading. In some experiments, RNA controls were added only before the RT-PCR reaction. Amplification protocols used the ready-to-go RT-PCR beads (Amersham Pharmacia Biotech) as described [7]. In several experiments, amplifications were performed in the presence of ^{32}P -labeled dCTP. The reaction mixtures after amplification were treated with RNase A and the products were resolved on a 5% non-denaturing acrylamide gel, unless stated otherwise. The gel was stained with ethidium bromide and photographed under UV light. When amplified products were ^{32}P -labeled,

products were quantified on an InstantImager (Canberra-Packard) or a Storm PhosphorImager (Amersham Biosciences) and then exposed on film.

Immunoprecipitation assay

Plasmids p7-AdA(-.-) and (a.a) were linearized with BsmAI and transcribed with T3 RNA polymerase to generate uniformly labeled RNA. Briefly, splicing reactions containing 10^5 cpm of ^{32}P -labeled transcripts, were incubated in a HeLa nuclear extract for 30 min at 30°C , either in the presence or the absence of oligonucleotides. Reactions were then placed on ice and incubated for 30 min with 1 μl of an antibody against hnRNP A1/A2 [45]. Fifty μl of protein A Sepharose (5 mg, Amersham Pharmacia Biotech) were added and the mixture was incubated for 15 min. After 3 washes with 1 ml of NET-2 buffer (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.5 mM DTT; and 0.05% Nonidet P-40), samples were PCA extracted and ethanol precipitated. The RNA recovered was fractionated on a denaturing 5% polyacrylamide gel.

Gel shift and RNase H protection assays

The gel-shift assay with recombinant hnRNP proteins was carried out as described [43]. ΔU1 HeLa extracts were produced by addition of the 2'O-Me oligo CCUGCCAGGUAAGUA complementary to the 5' end of U1 snRNA [46]. The oligonucleotide-mediated RNase H protection assay was conducted as described [24].

Transfection and RNA analysis

COS-7 cells were cultured in Dulbecco's modified Eagle's medium, supplemented with 10% fetal bovine serum. Twenty-four hours prior to plasmid transfections, cells were seeded in 6-well plates (35 mm) at a density of 2.5×10^5 cells/well. At 50% confluence, the cells were transiently transfected with 2 μ g of the purified plasmid p44:V3 $\lambda\lambda\lambda\lambda\lambda$ [2] using Lipofectamine PLUS (Invitrogen). Twenty-four hours later, the cells were treated with different 2'-O-Me RNA oligos to a final concentration of 50 nM, and the cell culture was continued for 24 h.

Total RNA was prepared using TRIzol (Invitrogen) and treated with DNase I, according to the manufacturer's indications. Reverse transcription was performed with Omniscript RT (Quiagen) and the RT3 primer (GAAGGCACAGTCGAGGCTG), which anneals to the 3'-UTR of the pcDNA3 vector to avoid the interference of the endogenous CD44 mRNA. The reaction was carried out at 37°C for 60 min, stopped at 95°C for 5 min and ice-quenched, followed by PCR amplification in the presence of [α - 32 P]dCTP, and of oligo primers directed to CD44 exons 5 and 16 (AGTGAAAGGAGCAGCACTTCAGG and TCAGATCCATGAGTGGTATGGGAC, respectively). The amplification procedure was as follows: 95°C for 5 min, 35 cycles at 94°C for 30 sec, 56°C for 30 sec and 72°C for 30 sec; with a final extension at 72°C for 15 min. Reaction products were resolved by electrophoresis in a non-denaturing 5% polyacrylamide gel and quantified using the InstantImager system (Canberra-Packard).

Computational analysis

Human introns were from NCBI build 35.1. A total of 156525 introns of sizes greater than 330 bp were retained for further analysis. One hundred rounds of shuffling were carried out in selected portions (+11 to +150 and -41 to -180) of all introns to calculate random occurrence and standard deviations.

ACKNOWLEDGMENTS

We thank Aline Simoneau, Maryse Gendron and Johanne Toutant for the preparation of nuclear extracts. We thank Marco Blanchette, Stephen Hutchison, Aline Simoneau and Johanne Toutant for plasmids and the preparation of recombinant proteins. We are grateful to Doug Black for hnRNP F/H expression vectors and antibodies, and Gavin Screaton for p44:V3λλλλλλλ. This work was supported by a grant from the Canadian Institutes of Health Research (CIHR) to B.C. B.C. is a Canada Research Chair in Functional Genomics, and is a member of the Sherbrooke RNA/RNP group supported by the CIHR, the Université de Sherbrooke and the FCAR.

REFERENCES

1. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, et al. (2001) Initial sequencing and analysis of the human genome. *Nature* 409: 860-921.
2. Bell MV, Cowper AE, Lefranc MP, Bell JI, Screaton GR (1998) Influence of intron length on alternative splicing of CD44. *Mol Cell Biol* 18: 5930-5941.
3. Mayeda A, Krainer AR (1992) Regulation of alternative pre-mRNA splicing by hnRNP A1 and splicing factor SF2. *Cell* 68: 365-375.
4. Chabot B, LeBel C, Hutchison S, Nasim FH, Simard MJ (2003) Heterogeneous nuclear ribonucleoprotein particle A/B proteins and the control of alternative splicing of the mammalian heterogeneous nuclear ribonucleoprotein particle A1 pre-mRNA. In: Jeanteur Ph, editor. *Regulation of alternative splicing*. Berlin Heidelberg: Springer-Verlag. pp. 59-88.
5. Zhu J, Mayeda A, Krainer AR (2001) Exon identity established through differential antagonism between exonic splicing silencer-bound hnRNP A1 and enhancer-bound SR proteins. *Mol Cell* 8: 1351-1361.
6. Blanchette M, Chabot B (1999) Modulation of exon skipping by high-affinity hnRNP A1-binding sites and by intron elements that repress splice site utilization. *EMBO J* 18: 1939-1952.
7. Nasim FU, Hutchison S, Cordeau M, Chabot B (2002) High-affinity hnRNP A1 binding sites and duplex-forming inverted repeats have similar effects on 5' splice site

selection in support of a common looping out and repression mechanism. RNA 8: 1078-1089.

8. Nussinov R (1989) Conserved signals around the 5' splice sites in eukaryotic nuclear precursor mRNAs: G-runs are frequent in the introns and C in the exons near both 5' and 3' splice sites. J Biomol Struct Dyn 6: 985-1000.

9. Nussinov R (1988) Conserved quartets near 5' intron junctions in primate nuclear pre-mRNA. J Theor Biol 133: 73-84.

10. Nussinov R (1987) (A)GGG(A), (A)CCC(A) and other potential 3' splice signals in primate nuclear pre-mRNA sequences. Biochim Biophys Acta 910: 261-270.

11. Solovyev VV, Salamov AA, Lawrence CB (1994) Predicting internal exons by oligonucleotide composition and discriminant analysis of spliceable open reading frames. Nucleic Acids Res 22: 5156-5163.

12. Engelbrecht J, Knudsen S, Brunak S (1992) G+C-rich tract in 5' end of human introns. J Mol Biol 227: 108-113.

13. McCullough AJ, Berget SM (1997) G triplets located throughout a class of small vertebrate introns enforce intron borders and regulate splice site selection. Mol Cell Biol 17: 4562-4571.

14. Majewski J, Ott J (2002) Distribution and characterization of regulatory elements in the human genome. Genome Res 12: 1827-1836.

15. Yeo G, Hoon S, Venkatesh B, Burge CB (2004) Variation in sequence and organization of splicing regulatory elements in vertebrate genes. *Proc Natl Acad Sci U S A* 101: 15700-15705.
16. Han K, Yeo G, An P, Burge CB, Grabowski PJ (2005) A combinatorial code for splicing silencing: UAGG and GGGG motifs. *PLoS Biol* 3: e158.
17. Sirand-Pugnet P, Durosay P, Brody E, Marie J (1995) An intronic (A/U)GGG repeat enhances the splicing of an alternative intron of the chicken beta-tropomyosin pre-mRNA. *Nucleic Acids Res* 23: 3501-3507.
18. Carlo T, Sierra R, Berget SM (2000) A 5' splice site-proximal enhancer binds SF1 and activates exon bridging of a microexon. *Mol Cell Biol* 20: 3988-3995.
19. Carlo T, Sterner DA, Berget SM (1996) An intron splicing enhancer containing a G-rich repeat facilitates inclusion of a vertebrate micro-exon. *RNA* 2: 342-353.
20. Burd CG, Dreyfuss G (1994) RNA binding specificity of hnRNP A1: significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing. *EMBO J* 13: 1197-1204.
21. Min H, Chan RC, Black DL (1995) The generally expressed hnRNP F is involved in a neural-specific pre-mRNA splicing event. *Genes Dev* 9: 2659-2671.
22. Hastings ML, Wilson CM, Munroe SH (2001) A purine-rich intronic element enhances alternative splicing of thyroid hormone receptor mRNA. *RNA* 7: 859-874.

23. Caputi M, Zahler AM (2001) Determination of the RNA binding specificity of the heterogeneous nuclear ribonucleoprotein (hnRNP) H/H'/F/2H9 family. *J Biol Chem* 276: 43850-43859.
24. Chabot B, Blanchette M, Lapierre I, La Branche H (1997) An intron element modulating 5' splice site selection in the hnRNP A1 pre-mRNA interacts with hnRNP A1. *Mol Cell Biol* 17: 1776-1786.
25. LaBranche H, Dupuis S, Ben-David Y, Bani MR, Wellinger RJ, et al. (1998) Telomere elongation by hnRNP A1 and a derivative that interacts with telomeric repeats and telomerase. *Nat Genet* 19: 199-202.
26. McKay SJ, Cooke H (1992) hnRNP A2/B1 binds specifically to single stranded vertebrate telomeric repeat TTAGGGn. *Nucleic Acids Res* 20: 6461-6464.
27. Villemaire J, Dion I, Elela SA, Chabot B (2003) Reprogramming alternative pre-messenger RNA splicing through the use of protein-binding antisense oligonucleotides. *J Biol Chem* 278: 50031-50039.
28. Buvoli M, Cobianchi F, Biamonti G, Riva S (1990) Recombinant hnRNP protein A1 and its N-terminal domain show preferential affinity for oligodeoxynucleotides homologous to intron/exon acceptor sites. *Nucleic Acids Res* 18: 6595-6600.
29. Romano M, Marcucci R, Buratti E, Ayala YM, Sebastio G, et al. (2002) Regulation of 3' splice site selection in the 844ins68 polymorphism of the cystathionine Beta -synthase gene. *J Biol Chem* 277: 43821-43829.

30. Chou MY, Rooke N, Turck CW, Black DL (1999) hnRNP H is a component of a splicing enhancer complex that activates a c-src alternative exon in neuronal cells. *Mol Cell Biol* 19: 69-77.
31. Caputi M, Zahler AM (2002) SR proteins and hnRNP H regulate the splicing of the HIV-1 tev-specific exon 6D. *EMBO J* 21: 845-855.
32. Chen CD, Kobayashi R, Helfman DM (1999) Binding of hnRNP H to an exonic splicing silencer is involved in the regulation of alternative splicing of the rat beta-tropomyosin gene. *Genes Dev* 13: 593-606.
33. Buratti E, Baralle M, De Conti L, Baralle D, Romano M, et al. (2004) hnRNP H binding at the 5' splice site correlates with the pathological effect of two intronic mutations in the NF-1 and TSHbeta genes. *Nucleic Acids Res* 32: 4224-4236.
34. Garneau D, Revil T, Fiset JF, Chabot B (2005) hnRNP F/H proteins modulate the alternative splicing of the apoptotic mediator Bcl-x. *J Biol Chem* 280: 22641-22650.
35. McCullough AJ, Berget SM (2000) An intronic splicing enhancer binds U1 snRNPs to enhance splicing and select 5' splice sites. *Mol Cell Biol* 20: 9225-9235.
36. Cartegni L, Maconi M, Morandi E, Cobianchi F, Riva S, et al. (1996) hnRNP A1 selectively interacts through its Gly-rich domain with different RNA-binding proteins. *J Mol Biol* 259: 337-348.
37. Ding J, Hayashi MK, Zhang Y, Manche L, Krainer AR, et al. (1999) Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. *Genes Dev* 13: 1102-1115.

38. Fiset S, Chabot B (2001) hnRNP A1 may interact simultaneously with telomeric DNA and the human telomerase RNA in vitro. *Nucleic Acids Res* 29: 2268-2275.
39. Charpentier B, Rosbash M (1996) Intramolecular structure in yeast introns aids the early steps of in vitro spliceosome assembly. *RNA* 2: 509-522.
40. Howe KJ, Ares M, Jr. (1997) Intron self-complementarity enforces exon inclusion in a yeast pre-mRNA. *Proc Natl Acad Sci U S A* 94: 12467-12472.
41. Libri D, Stutz F, McCarthy T, Rosbash M (1995) RNA structural patterns and splicing: molecular basis for an RNA-based enhancer. *RNA* 1: 425-436.
42. Newman A (1987) Specific accessory sequences in *Saccharomyces cerevisiae* introns control assembly of pre-mRNAs into spliceosomes. *EMBO J* 6: 3833-3839.
43. Hutchison S, LeBel C, Blanchette M, Chabot B (2002) Distinct sets of adjacent heterogeneous nuclear ribonucleoprotein (hnRNP) A1/A2 binding sites control 5' splice site selection in the hnRNP A1 mRNA precursor. *J Biol Chem* 277: 29745-29752.
44. Dignam JD, Lebovitz RM, Roeder RG (1983) Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res* 11: 1475-1489.
45. Patry C, Bouchard L, Labrecque P, Gendron D, Lemieux B, et al. (2003) Small interfering RNA-mediated reduction in heterogeneous nuclear ribonucleoparticule A1/A2 proteins induces apoptosis in human cancer cells but not in normal mortal cell lines. *Cancer Res* 63: 7679-7688.

46. Barabino SM, Blencowe BJ, Ryder U, Sproat BS, Lamond AI (1990) Targeted snRNP depletion reveals an additional role for mammalian U1 snRNP in spliceosome assembly. Cell 63: 293-302.

CHAPITRE 2

Fisette J-F., Toutant J., Dugré-Brisson S., DesGroseillers L. and Chabot B. hnRNP A1 and hnRNP H can collaborate to modulate 5' splice site selection, (article soumis dans *RNA*).

RÉSUMÉ

Les protéines hnRNP A1 et hnRNP H sont impliquées dans plusieurs événements d'épissage alternatif. Un mécanisme proposé afin d'expliquer ces effets implique une interaction entre molécules hnRNP A1 liées à l'ARN qui permet un rapprochement de sites d'épissage distants. Notre étude montre que des sites de liaison pour hnRNP H peuvent reproduire l'activité des motifs de liaison pour hnRNP A1. De plus, des combinaisons hétérotypiques de sites de liaison pour ces molécules peuvent aussi moduler l'épissage. La portion en C-terminal de hnRNP H, correspondant aux deux domaines riches en résidus glycines (GRD) et au quasi-RRM3, est importante pour cette activité d'épissage. La substitution de cette région par le GRD de hnRNP A1 préserve l'activité. Finalement, nous avons utilisé la technologie BRET (« bioluminescence resonance energy transfer ») afin de montrer que des protéines hnRNP H et hnRNP A1 peuvent former des partenaires d'interaction homotypiques et hétérotypiques dans les cellules vivantes. Nos résultats suggèrent que des interactions entre molécules hnRNP liées à l'ARN peuvent moduler la sélection des sites d'épissage.

J'ai contribué à la réalisation de toutes les expériences. Plus précisément, j'ai effectué toutes les expériences d'épissage *in vitro*. J'ai aussi effectué toutes les expériences de liaison des protéines à l'ARN. J'ai fait les expériences de BRET sous la supervision de Samuel Dugré-Brisson. J'ai contribué à l'écriture de cet article.

hnRNP A1 and hnRNP H can collaborate to modulate 5' splice site selection

JEAN-FRANÇOIS FISETTE¹, JOHANNE TOUTANT¹, SAMUEL DUGRÉ-BRISSON²,
LUC DESGROSEILLERS² and BENOIT CHABOT^{1,3}

¹Département de microbiologie et d'infectiologie, Faculté de Médecine et des Sciences de la Santé, Université de Sherbrooke, Sherbrooke, Québec, Canada. J1H 5N4, ²Département de biochimie, Université de Montréal, Montréal, Québec, Canada, H3C 3J7

Running title: Splicing modulation by hnRNP H and hnRNP A1

³ Corresponding author: email: Benoit.Chabot@USherbrooke.ca

ABSTRACT

The mammalian proteins hnRNP A1 and hnRNP H control many splicing decisions in viral and cellular primary transcripts. To explain some of these activities, we have proposed that self-interactions between bound proteins create an RNA loop that represses internal splice sites while simultaneously activating the external sites that are brought in closer proximity. Here we show that a variety of hnRNP H binding sites can affect 5' splice site selection. The addition of two sets of hnRNP H sites in a model pre-mRNA modulates 5' splice site selection in a synergistic manner, consistent with the looping model. Importantly, binding sites for hnRNP A1 and H on the same pre-mRNA can similarly collaborate to modulate 5' splice site selection. The C-terminal portion of hnRNP H that contains the glycine rich domains (GRD) is essential for splicing activity, and it can be functionally replaced by the GRD of hnRNP A1. Finally, we used the Bioluminescence Resonance Energy Transfer (BRET) technology to document the existence of homotypic and heterotypic interactions involving hnRNP H and hnRNP A1 in live cells. Overall, our study suggests that interactions between different hnRNP proteins bound to distinct locations on a pre-mRNA can change its conformation to affect splicing decisions.

INTRODUCTION

As much as 74% of human pre-messenger RNAs (pre-mRNAs) may be alternatively spliced (Johnson et al. 2003). Alternative splicing represents an important mechanism to increase proteome diversity, hence contributing to the biological complexity of mammals (Graveley 2001). Appropriate selection of alternative splice sites is critical to produce proteins with specific functional attributes. The orchestration of this multitude of splicing decisions necessitates RNA binding proteins of which members of the SR and hnRNP protein families are the best characterized.

SR proteins play a role both in constitutive and alternative splicing (Graveley 2000). This family of proteins frequently enforces exon inclusion by binding to exonic splicing enhancers to stimulate the binding of generic splicing factors including U2AF, U2 snRNP and U1 snRNP (Lavigne et al. 1993; Kohtz et al. 1994; Wang et al. 1995; Zuo and Maniatis 1996; Graveley et al. 2001). However, SR proteins can also have negative impacts on splicing. This is the case in the hnRNP A1 pre-mRNA where SRp30c can repress the 3' splice site of exon 8 (Simard and Chabot 2002; Paradis et al. 2007) and in the adenovirus L1 unit where ASF/SF2 can prevent the recruitment of U2 snRNP (Kanopka et al. 1996).

On the other hand, hnRNP proteins have been regarded mostly as splicing repressors. Indeed, hnRNP A1 can interfere with the binding of the SR protein SC35 preventing it to stimulate the inclusion of HIV *tat* exon 2 (Zahler et al. 2004). Moreover, hnRNP A1 can inhibit the inclusion of an alternative exon in human *Ras*, *SMN2*, β -*tropomyosin* and *c-src* genes (Guil et al. 2003; Kashima and Manley 2003; Rooke et al.

2003; Expert-Bezancon et al. 2004). Likewise, hnRNP H can repress exon inclusion in several genes: HIV *tat*, α -tropomyosin, *c-src* and *fibroblast growth factor receptor 2* (Min et al. 1995; Chou et al. 1999; Jacquenet et al. 2001; Crawford and Patton 2006; Mauger et al. 2008). However, positive roles for hnRNP proteins have also been observed. Intronic binding sites for hnRNP A1 and hnRNP H can stimulate *in vitro* splicing (Martinez-Contreras et al. 2006). hnRNP H and A1 can also have opposite roles as was documented in the CI alternative cassette exon of the glutamate NMDA R1 receptor (GRIN1) transcript (Han et al. 2005). In this case, hnRNP A1 was shown to mediate silencing while hnRNP H antagonized this activity. In a recent analysis using RNA interference and a RT-PCR screening platform, we examined the role of the major hnRNP proteins in the splicing of apoptotic genes (Venables et al. 2008). Our study suggests that approximately 10% of all alternative splicing events are controlled by either hnRNP A1 or hnRNP H (Venables et al. 2008). Moreover, the effects obtained by knocking-down these and other hnRNP proteins suggest that hnRNP proteins can promote a mixture of exon inclusion/exclusion events, and that the identity of the target genes varies considerably in different cellular contexts.

The enhancing or silencing activity of hnRNP A1 and hnRNP H in splicing vary depending on the position of their respective binding sites relative to the splicing signals. For example, in HIV-1 *tev* exon 6D, an hnRNP H binding site located 66 nucleotides from the 5' splice site promotes U1 snRNP binding (Caputi and Zahler 2002), while in the *NF-1* and *TSH β* genes, binding of hnRNP H at the 5' splice site restricts U1 snRNP binding (Buratti et al. 2004). hnRNP A1 binding sites can stimulate splicing between an external pair of splice sites while decreasing the use of an intervening one (Nasim et al. 2002;

Martinez-Contreras et al. 2006). To explain the role of hnRNP A1 and H binding sites in alternative splicing we have proposed the looping out model which posits that bound hnRNP proteins self-interact to loop out and repress internal splice sites, while simultaneously juxtaposing and stimulating the use of an external pair of splice sites (Chabot et al. 1997; Blanchette and Chabot 1999; Martinez-Contreras et al. 2006). Consistent with this model, inverted repeats can mimic the impact of A1 binding sites (Nasim et al. 2002). The enrichment of putative binding sites for hnRNP A1 and hnRNP H proteins (GGG motifs) near both ends of introns is also consistent with a role for this mechanism in generic splicing (Nussinov 1987; Nussinov 1988; Nussinov 1989; Engelbrecht et al. 1992; Solovyev et al. 1994; McCullough and Berget 1997; Majewski and Ott 2002; Yeo et al. 2004).

hnRNP A1 and hnRNP H are structurally related proteins. hnRNP A1 contains two N-terminal RNA recognition motifs (RRM1 and RRM2), while the equivalent region in hnRNP H hosts two motifs that contain small similarity with the classical RRM but can bind RNA and were denoted quasi-RRMs (qRRM1 and qRRM2) (Honore et al. 1995). The C-terminal half of hnRNP A1 contains a glycine-rich domain that is required for self-interaction (Cartegni et al. 1996) and is essential for the activity of A1 in splicing (Mayeda et al. 1994). hnRNP H contains two glycine-rich domains interrupted by a qRRM (qRRM3). In contrast to the situation for hnRNP A1, the function of the individual domains of hnRNP H has not yet been investigated. Given their structural similarity and their equivalent impact on splicing, it was of interest to determine if hnRNP A1 and hnRNP H could collaborate to enforce a specific splicing event. Our *in vitro* analysis

indicates that they do and their very close proximity in live cells suggests a direct interaction between hnRNP A1 and hnRNP H.

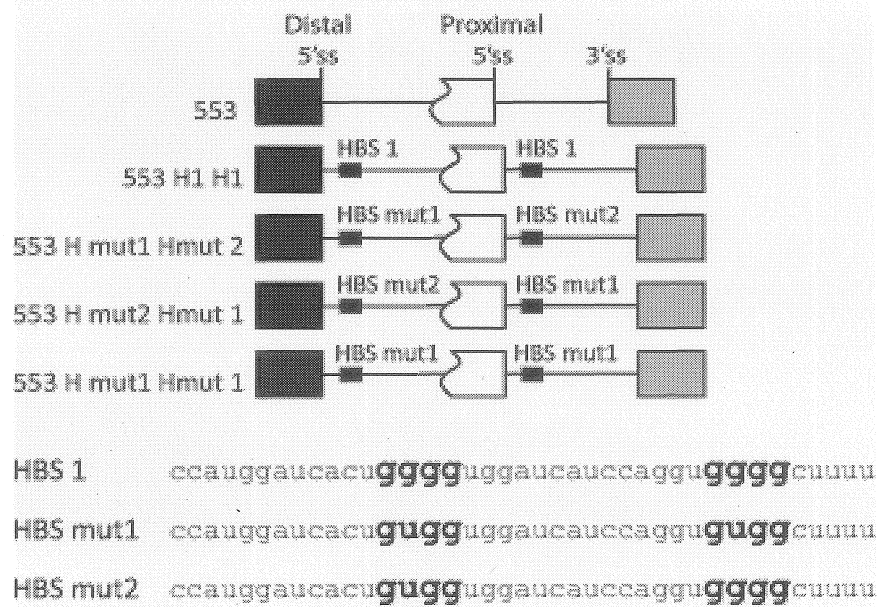
RESULTS

hnRNP H binding sites modulate 5' splice site selection

Previously, we have shown that inserting a hnRNP H binding site downstream of two competing 5' splice sites stimulates splicing to the most upstream site *in vitro* (Martinez-Contreras et al. 2006). The hnRNP H binding site used in our previous study was from the cystathionine β -synthase gene where it regulates 3' splice site selection (Romano et al. 2002). This site, that we now call HBS1, is bound by hnRNP H but not by hnRNP A1 (Martinez-Contreras et al. 2006). There are two G quadruplets present in HBS1 (Fig. 1A, bottom panel). Here we show that both G motifs contribute to the activity of this element since mutating either one or both G quadruplets reduced distal 5' splice site utilization (Fig. 1B, compare lanes 4, 5 and 6 with lane 3). Thus, the activity of HBS1 requires the integrity of both G quadruplets.

We then tested the activity of two different binding sites for hnRNP H. HBS2 contains a GGGG and a GGG motifs (Fig. 2A) and functions as a negative regulator in the splicing of the Rous sarcoma virus (Fogel and McNally 2000). HBS3 contains a single GGGGG sequence (Fig. 2A) and it is part of a polyadenylation signal in the *RATCRP2A* gene (Arhin et al. 2002). Since GGG represents the core sequence found in many binding sites for hnRNP H (Caputi and Zahler 2001) and is also part of the high-affinity binding site for hnRNP A1 (Burd and Dreyfuss 1994), we set out to assess if HBS2 and HBS3 were bound specifically by hnRNP H. Using a filter binding assay, we show that HBS2 and HBS3 are bound more efficiently by recombinant hnRNP H than by hnRNP A1 (Fig. 2B). In contrast, the hnRNP A1 binding site CE1a, which exists in the

A



B

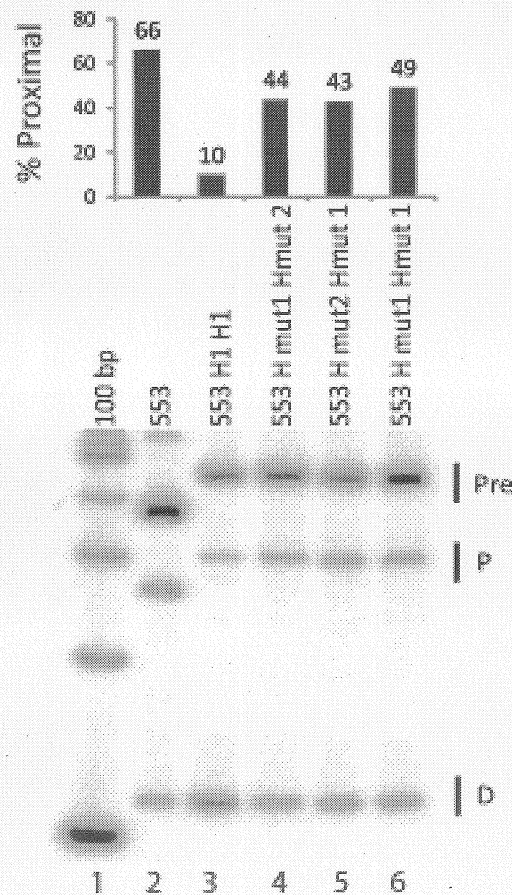


Figure 1. Different G runs affect distal 5' splice site utilization. (A) Schematic representation of pre-mRNAs containing wild type (HBS1) or mutated hnRNP H binding sites (HBSmut1 and HBSmut2). (B) The pre-mRNAs were incubated in HeLa nuclear extract at 30°C for 2 h and splicing products were amplified by RT-PCR. Products corresponding to distal 5' splice site utilization (D), proximal 5' splice site usage (P) or pre-mRNAs (Pre) are indicated. 100 bp indicates molecular weight markers. Each splicing reaction was performed in triplicate and the percentage of proximal products relative to the sum of the proximal and distal products were calculated. The *p* value for the difference between the wild type 553 H1H1 and each mutant was smaller than 0.05.

hnRNP A1 pre-mRNA (Chabot et al. 1997; Hutchison et al. 2002), is bound by hnRNP A1 but not H (Fig. 2B). Next, we tested if HBS2 and HBS3 could modulate alternative 5' splice site utilization in the same manner as HBS1 and CE1a. For this we used the 553 model pre-mRNA, which contains two 5' splice sites competing for a single 3' splice site (Fig. 2A). Pre-mRNAs were incubated in HeLa nuclear extracts, and the relative use of the competing 5' splice sites was assessed by performing a RT-PCR assay on the *in vitro* mRNA splicing products. In the absence of hnRNP A1 binding sites, the 553 pre-mRNA is spliced predominantly to the proximal 5' splice site (Fig. 2C and 2D, lane 1). Positioning a HBS2 site downstream of either the proximal or the distal 5' splice site moderately stimulated distal 5' splice site utilization (Fig. 2C, compare lanes 2 and 3 with lane 1). Inserting HBS2 at both positions promoted a stronger shift (lane 4). When single HBS3 sites were inserted, activity was observed only with the site at the upstream position; however, insertion of another site downstream significantly improved distal 5' splice site use (Fig. 2D, lanes 2-4).

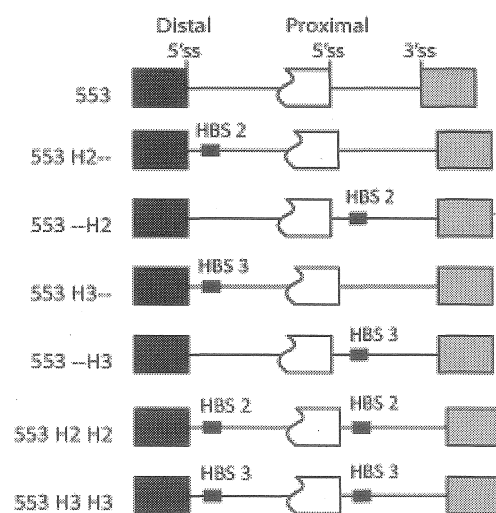
If the looping out model is valid, the activity of HBS should be observed only when HBS are present at both positions on our model pre-mRNAs. This was the case

A

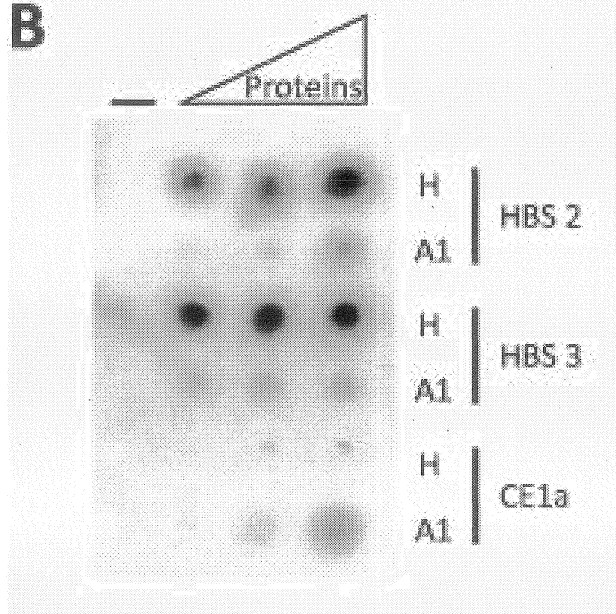
HBS 2 *cgaaucgacaaa***ggggg**aggaagu**ggg**agaaa

HBS 3 *ccauggguuu***ggggg**caguaguugg

CE1a *uaccuu***uagagu**aggc



B



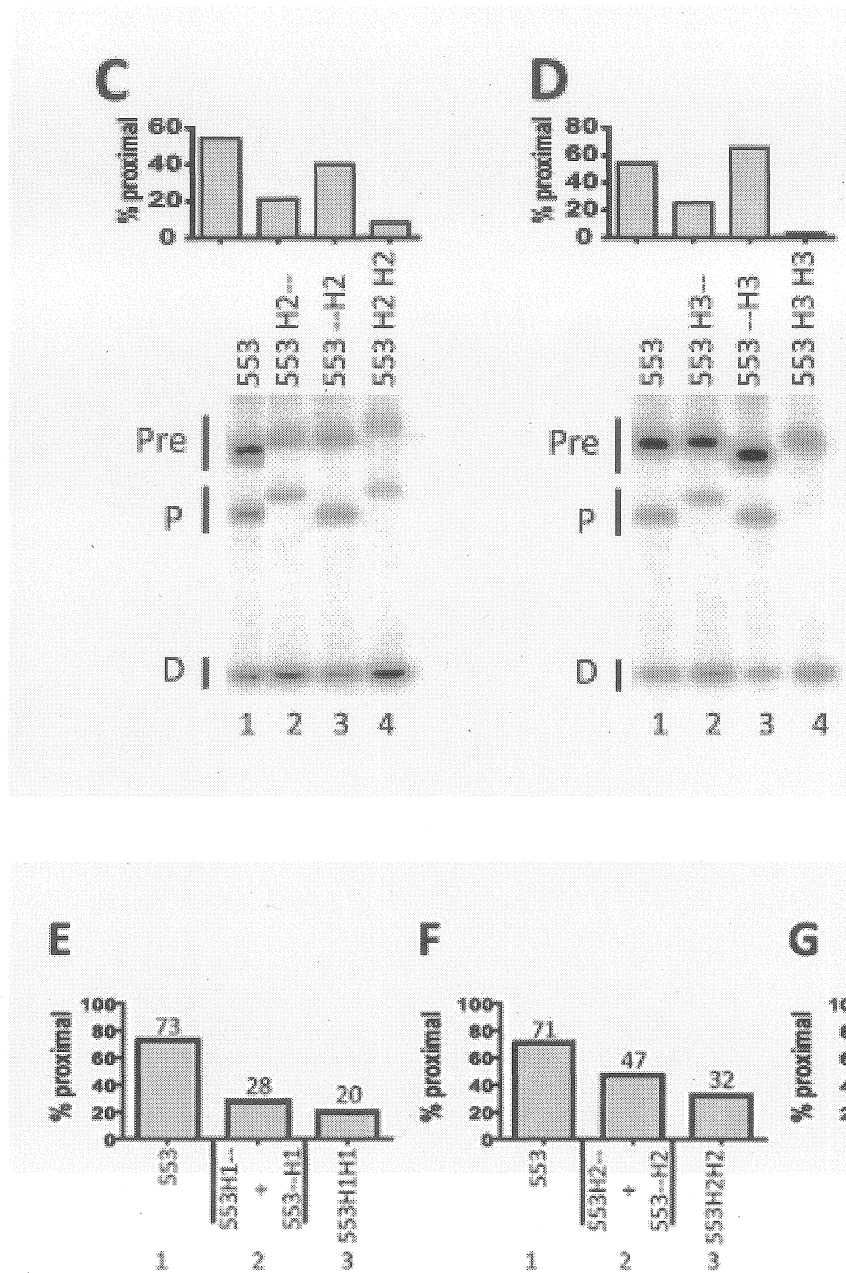
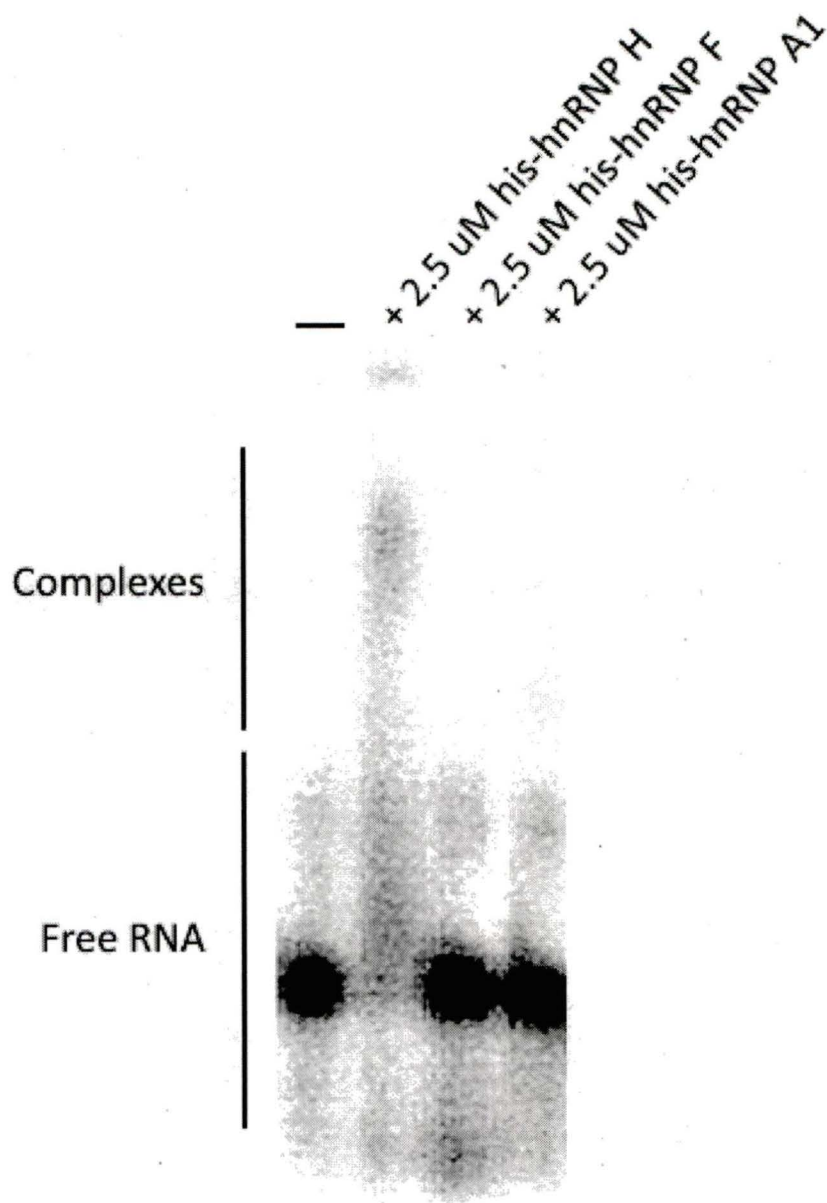


Figure 2. Binding sites for hnRNP H can modulate splice site selection. (A) Sequence of the binding sites HBS2, HBS3 and CE1a (left). Structure of the pre-mRNAs containing hnRNP H binding sites (right). (B) hnRNP H binds more efficiently to HBS2 and HBS3 compared with hnRNP A1. Filter binding assay using 32 P-labeled HBS2 or HBS3 or CE1a RNAs. His-tagged hnRNP H or A1 proteins were used at 0, 0.5, 1 and 2 μ M. (C and D) *In vitro* splicing assays and RT-PCR amplification. The distal (D), and proximal (P) products as well as the pre-mRNAs (Pre) are indicated. For each reaction, the percentage of proximal product is indicated in the graph. (E, F, and G) *In vitro* splicing assays in the presence of 10 pmoles of DNA oligonucleotide HUMMK (CCATGGTTTGGGAGTGGGAAGGTGGGGAG) containing hnRNP H binding sites.

Three different splicing reactions, using pre-mRNAs containing HBS 1 (E), HBS 2 (F) or HBS 3 (G), were performed. The sum of the activity of one binding site for hnRNP H at either positions (lane 2) was inferior than the activity obtained with a combination of hnRNP H binding sites (lane 3).

when measuring the stimulatory impact of single and double HBS inserted at the ends of a long intron (Martinez-Contreras et al. 2006). Although this is the configuration that is the most active with the 553 pre-mRNA used here, molecules containing a single HBS at almost any of the two positions were also moderately active. As discussed previously for hnRNP A1 (Martinez-Contreras et al. 2006), we attribute the activity of single HBS to the presence of weaker binding sites for hnRNP H in other portions of the pre-mRNA that may provide partners for H bound to the higher affinity site. Since support for the looping model would increase if we can show that the impact of two strong HBS is more than sum of the individual ones, we tried to reduce the contribution of weaker sites by carrying out splicing reactions in the presence of small quantities of a DNA oligo (HUMMK) containing hnRNP H binding sites taken from the polyadenylation signal in the *HUMMK* gene (Arhin et al. 2002). This DNA oligo (CCATGGTTTGGGAGTGGGAAGGTGGGGAG) is bound by recombinant hnRNP H but not A1 (Supp. Fig 1). We performed *in vitro* splicing assays on pre-mRNAs carrying HBS1, HBS2 or HBS3 in the presence of HUMMK. For each experiment, we compared the impact of having H binding sites at both positions (Fig. 2E-G, lane 3) with the sum of the impact of having HBS at individual positions (lane 2). In all three cases, the impact of two HBS on the same pre-mRNA was stronger than the additive impact of the individual HBS. Thus, the synergistic effect produced by hnRNP H proteins bound to two HBS on the same pre-mRNA is consistent with the looping model.

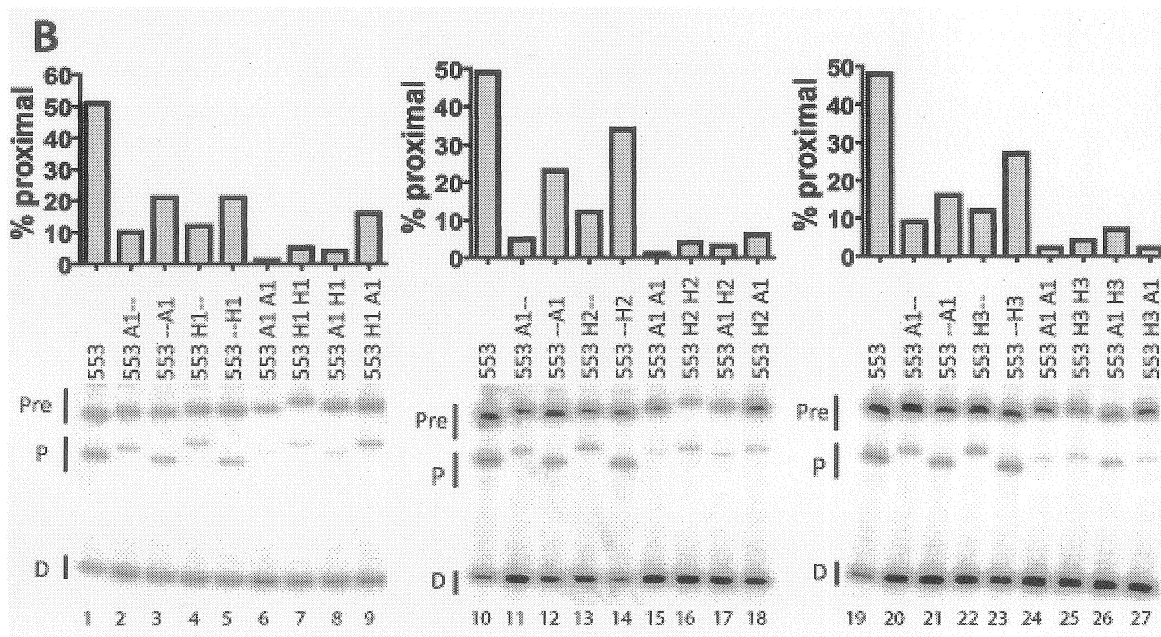
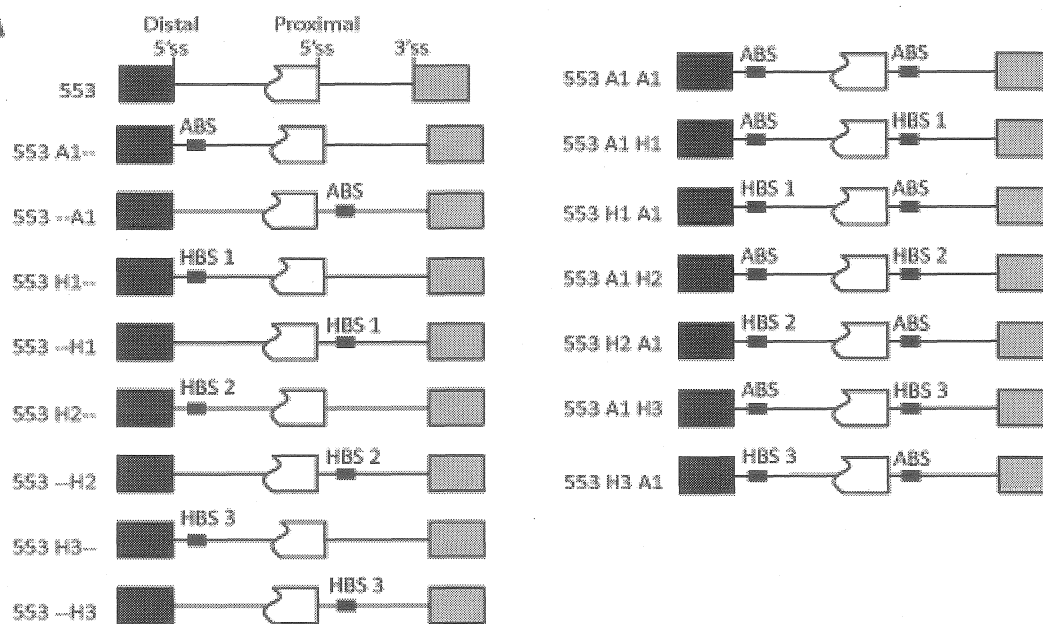


Supp. figure 1. hnRNP H but not hnRNP A1 and hnRNP F binds HUMMK. Binding of recombinant hnRNP H, F and A1 to ^{32}P 5' end-labeled HUMMK was monitored by gel shift assay. 0 and 2.5 μM of recombinant proteins were tested.

Combining hnRNP A1 and hnRNP H binding sites

Our *in vitro* splicing assays show that having ABS or HBS at two separate positions on a pre-mRNA can have a strong effect on 5' splice site selection. Next, we asked if HBS and ABS could cooperate on the same pre-mRNA to stimulate distal 5' splice site usage. We constructed model pre-mRNAs carrying ABS combined with different HBS in the 553 pre-mRNA backbone (Fig. 3A). Figure 3B confirms the activity of one and two ABS (lanes 2, 3, 6, 11, 12, 15, 20, 21 and 24) as well as single and dual HBS (lanes 4, 5, 7, 13, 14, 16, 22, 23 and 25). Notably, the ABS combination always worked slightly better than the HBS combinations. Combining one ABS with any of the HBS in different configurations always stimulated distal 5' splice site usage more than one ABS or one HBS (Fig. 3B, lanes 8, 17, 18, 26, 27), except for the HBS1-ABS configuration (lane 9). The HBS3-ABS set (lane 27) was as active as ABS-ABS (lane 24). Thus, elements bound by A1 and H can function on the same pre-mRNA to modulate 5' splice site selection. In these experiments, the effects of the majority of the combinations corresponded to the sum of the activity of individual sites. To minimize the impact of weaker binding sites for hnRNP H and A1 that likely exist on the pre-mRNA, we repeated the splicing assays using 553 A1-H2 and 553 H3-A1 pre-mRNAs in the presence of a small amount of oligo HUMMK and oligonucleotide TS10 (Dallaire et al. 2000) carrying 10 copies of binding sites for hnRNP A1 ([TTAGGG]₁₀). In all cases, the activity of the combined sites were superior than the sum of the effect of the individual sites (Fig. 3C and 3D). These results suggest that bound hnRNP A1 and H can function in synergy.

A



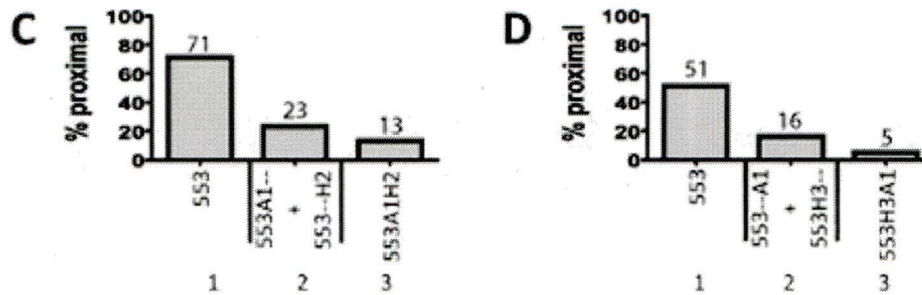


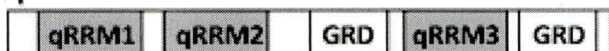
Figure 3. HBS and ABS function together to modulate distal 5' splice site usage. (A) Structure of the 553 splicing units containing HBS1, HBS2, HBS3 and ABS. (B) The pre-mRNAs were incubated in HeLa nuclear extract. Pre-mRNAs and splicing products were amplified by RT-PCR. For each reaction, the percentage of proximal product is indicated in the graph. (C-D) *In vitro* splicing assays in the presence of 2 pmoles of DNA oligo HUMMK containing hnRNP H binding sites (CCATGGTTTGGGAGTGGGAAGGTGGGGAG) and 2 pmoles of DNA oligo TS10 carrying hnRNP A1 binding sites ((TTAGGG)₁₀). Two different splicing reactions, using pre-mRNAs containing ABS and HBS 2 (C), or ABS and HBS 3 (D), were performed. The sum of the activity of one binding site for hnRNP H or hnRNP A1 at both positions was less than the activity of a combination of hnRNP H and hnRNP A1 binding sites.

Modularity of the hnRNP A1 and H domains

The A1-mediated effects on 5' splice site selection require the glycine-rich domain of hnRNP A1 (Mayeda et al. 1994; Blanchette and Chabot 1999). Given the equivalent activity of ABS and HBS and the similarity in domains between hnRNP H and A1 proteins, the glycine-rich domains (GRD) of H and A1 may be interchangeable. We tested this hypothesis by constructing chimeric proteins (Fig. 4A). First, we assessed the binding capacity of these proteins using a gel shift assay. The GRD-qRRM3-GRD portion of hnRNP H was not essential for RNA binding (Hs; Fig. 4B, lanes 8 and 9). Moreover, linking the qRRM1-qRRM2 portion of hnRNP H with the GRD of A1 (HA1) preserved binding (lanes 12 and 13). In contrast, the hnRNP F protein, which is 78% identical to hnRNP H (Caputi and Zahler 2001), did not bind to HBS1 (lanes 4-5).

A

F



H



Fs



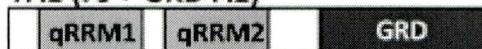
Hs



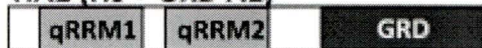
A1



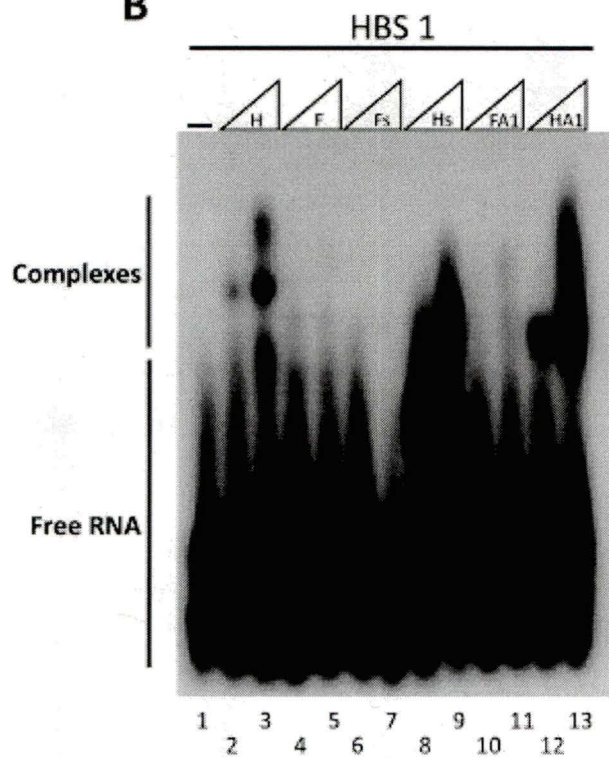
FA1 (Fs + GRD A1)



HA1 (Hs + GRD A1)



B



extracts (Fig. 4C). In such conditions, the percentage of proximal products derived from the 553 H1H1 pre-mRNA increases substantially and is strongly reduced by the addition of recombinant his-tagged hnRNP H protein (Fig. 4D, lanes 2-4). Supplementation with the truncated Hs protein that lacks the GRD-qRRM3-GRD region had no effect (lanes 9 and 10), as hnRNP F (lanes 5-6), its truncated derivative (lanes 7-8) or the chimeric FA1 protein (lanes 11-12). Remarkably, the HA1 protein restored distal 5' splice site selection to a level that was even superior to the level obtained with hnRNP H (Fig. 4C, compare lanes 13-14 with lanes 3-4), possibly reflecting its higher binding affinity for HBS1 or its improved capacity at forming protein-protein interactions. Thus, the GRD domain of hnRNP A1 can functionally substitute for the GRD-qRRM3-GRD domain of hnRNP H in a 5' splice site selection assay *in vitro*.

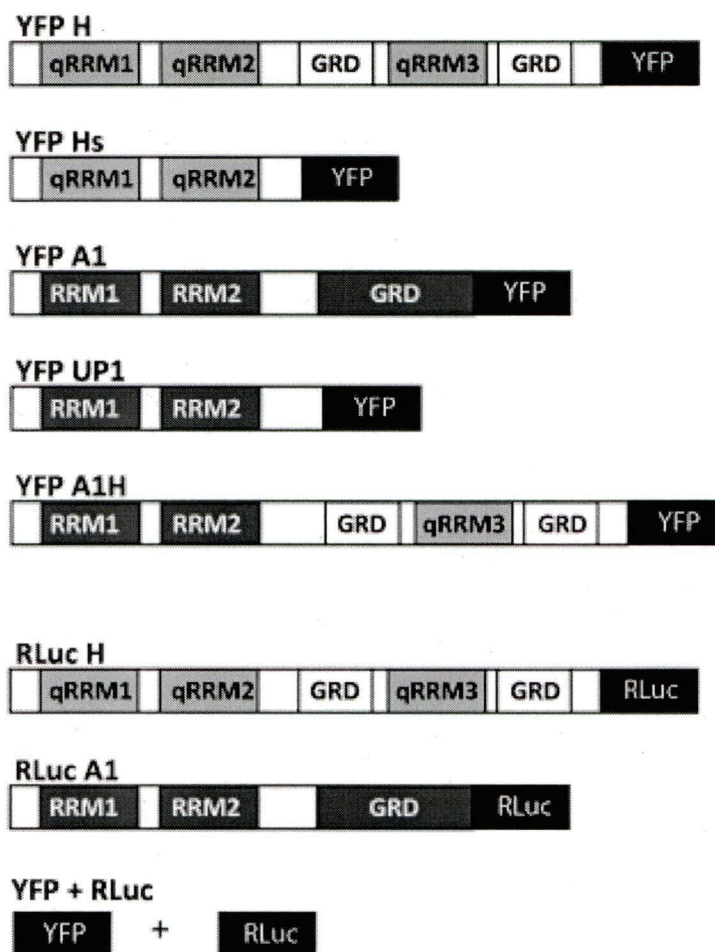
Interaction between hnRNP A1 and hnRNP H in live cells

Our *in vitro* splicing results suggest that hnRNP A1 and hnRNP H proteins can interact to modulate splice site selection. Distribution of G-triplets in the genome shows a bias toward enrichment in introns (Nussinov 1987; Nussinov 1988; Nussinov 1989; Engelbrecht et al. 1992; Solovyev et al. 1994; McCullough and Berget 1997; Majewski and Ott 2002; Yeo et al. 2004; Han et al. 2005). This large quantity of GGG triplets indicates a multitude of putative binding sites for A1 and H and, hence, possibilities of interaction between these proteins must be enormous. A recent study using TAP-tagged A1 and H revealed that at least a portion of these proteins could exist in the same complexes (Jeronimo et al. 2007). However, their spatial relationship and proximity were not addressed in this experiment. To examine whether hnRNP A1 and hnRNP H can interact with one another in a cell, we used the Bioluminescence Resonance Energy

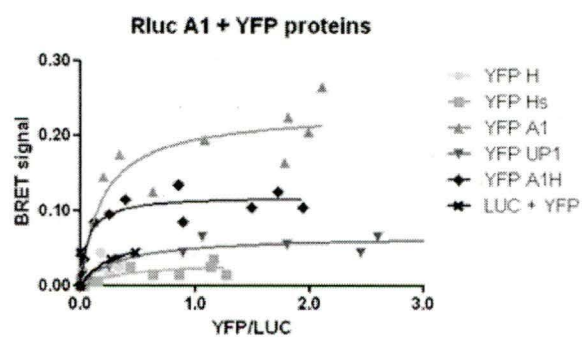
Transfer (BRET) technology. BRET is based on energy transfer from a donor (*Renilla* luciferase (*Rluc*)) to an acceptor molecule (YFP) and occurs when the donor/acceptor pair is within 50-100 Å (Xu et al. 1999), which is the typical distance separating the centers of two interacting proteins. Proximity between hnRNP H and hnRNP A1 proteins were tested in EcR 293 cells. An equivalent quantity of plasmids programmed to express *Rluc*-tagged proteins was co-transfected with an increasing amount of plasmids designed to express YFP-tagged proteins (Fig. 5A). The BRET signal is calculated as the ratio of the YFP excited light over luciferase emission. Because of variations in the expression profile of each plasmid, the BRET signal is plotted relative to total YFP activity/*Rluc* activity, which essentially normalizes YFP expression relative to the *Rluc* signal in each transfected sample. We first used the BRET assay to confirm the interaction between hnRNP A1 proteins since A1 can self-interact *in vitro* (Cartegni et al. 1996). A strong BRET signal was obtained when *Rluc* A1 was co-expressed in EcR 293 cells with increasing amounts of YFP A1 (Fig. 5B and 5C). As expected for a negative control, a weak signal was obtained when *Rluc* and YFP were co-expressed (Fig. 5B and C). Likewise, a similarly low BRET signal was obtained when the glycine-rich domain (GRD) of A1 was removed (YFP UP1), indicating that this interaction depends on the GRD domain (Fig. 5B and 5C).

To test whether a similar interaction can occur between hnRNP A1 and hnRNP H, we co-expressed *Rluc* A1 with YFP H. In this case, the BRET signal was low but this

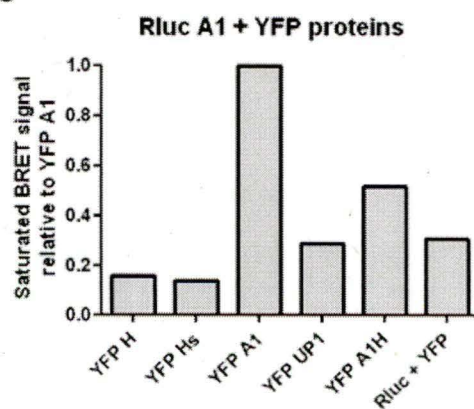
A



B



C



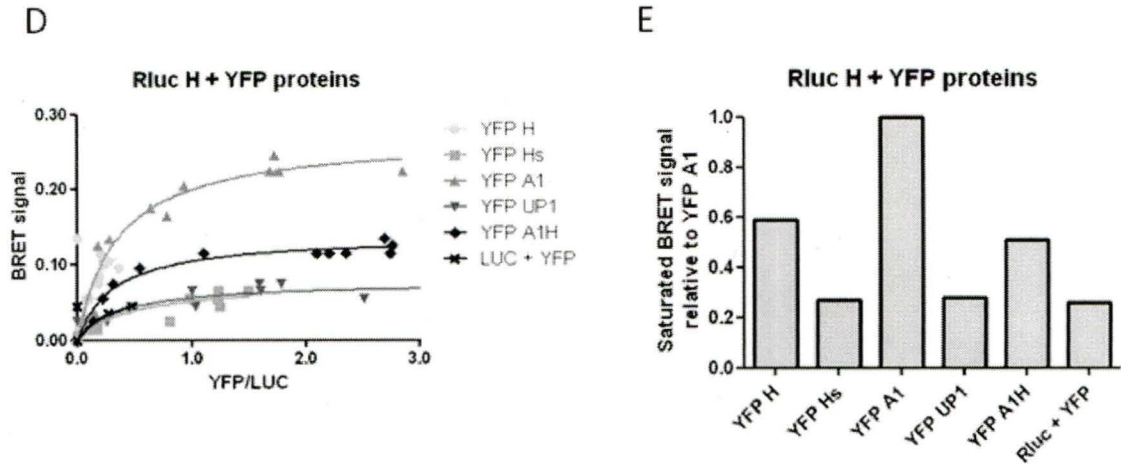


Figure 5. hnRNP H and hnRNP A1 can interact in live cells. (A) Structure of the YFP- and Rluc-tagged proteins used for BRET. (B and D) EcR 293 cells were transfected with 10 ng of plasmid expressing the Rluc version with increasing amounts (50, 100, 250 or 500 ng) of YFP-tagged plasmids. A control expressing only Rluc with an increasing amount of plasmids expressing only YFP was used (YFP + Rluc). Cells were collected 36 h post-transfection and the expression level of the YFP fused proteins was determined by fluorescence. The BRET signal was calculated as the fluorescence/luminescence after addition of 5 μ M coelenterazine h. The BRET signals were plotted relative to YFP/Rluc. (C and E) Plots showing the maximal saturated BRET signal for distinct combinations. The signals were normalized relative to YFP A1.

could be due to the low expression of YFP H (Fig. 5B). Because a BRET signal often depends on the configuration of the interaction and the design of the interacting proteins (Pfleger et al. 2006), we tested the reciprocal Rluc H design with YFP A1. Following expression in EcR 293 cells, strong BRET signals were obtained with Rluc H co-expressed with YFP A1 (Fig. 5D and 5E). The result obtained with YFP UP1 indicates that this interaction is dependent on the GRD of A1. Our results therefore indicate that hnRNP H can interact with hnRNP A1 in live cells. Likewise, a strong BRET signal was obtained when Rluc H was co-expressed with YFP H (Fig. 5D and 5E). The intensity of the signal dropped considerably when the GRD-qRRM3-GRD domain of H was removed

(YFP Hs; Fig. 5D and 5E), indicating that the hnRNP H self-interaction is dependent on this region. Overall, the A1-A1 and A1-H interactions were stronger than the H-H interaction. This conclusion was confirmed by substituting the GRD portion of A1 for that of H (YFP A1H); co-expressing *Rluc* H and YFP A1H produced a BRET signal that was in the same order than with YFP H. Overall, our results indicate that homotypic A1-A1 and H-H interactions occur in EcR 293 cells and that they require the GRD-containing portions of these proteins. Moreover and most importantly, heterotypic interactions between hnRNP H and hnRNP A1 can also occur in live cells.

DISCUSSION

Our study has investigated the role of hnRNP H binding sites as modulators of 5' splice site selection. hnRNP H binding sites containing GGG, GGGG or GGGGG in different sequence contexts demonstrated a similar ability at promoting distal 5' splice site selection using a model pre-mRNA in HeLa extracts. The essential role of G runs was also demonstrated by showing that mutating one G in one triplet was sufficient to abrogate its activity in 5' splice site selection.

The mechanisms by which hnRNP H bound to G-runs control splice site selection can vary. When G runs exist very near splice sites, the binding of hnRNP H can sterically interfere with splice site recognition (Jacquenet et al. 2001). In contrast, the presence of hnRNP H binding sites 30 nt downstream of each of two competing 5' splice sites shifted the ratio of use in favor of the distal 5' splice site. Because this shift was not accompanied by a change in the initial binding of U1 snRNP (Martinez-Contreras et al. 2006), we proposed that the mechanism of action in this case was similar to the model that we had proposed previously for hnRNP A1 (Blanchette and Chabot 1999; Nasim et al. 2002; Chabot et al. 2003), namely that bound hnRNP H proteins would self-interact to loop out and repress the intervening 5' splice site while simultaneously bringing in close proximity the distal 5' and the 3' splice sites. This model implies that an hnRNP H binding sites at only one position should not work. While this behavior was observed in one pre-mRNA substrate previously (Martinez-Contreras et al. 2006), a single hnRNP H binding site at only one position can confer some activity in the 553 pre-mRNA used here. Our preferred explanation for this behavior is that this added site can function in combination with other weaker hnRNP H sites on the pre-mRNA. Nevertheless, one prediction from our model is

that the addition of strong hnRNP H binding sites at the two positions should work better than the sum of the activity of the hnRNP H binding sites at individual positions. Because this is exactly what we observed for each of the three different hnRNP H binding sites tested, these results provide further support in favor of the looping out model of action for hnRNP H.

Having obtained strong indications that hnRNP H binding sites can function in a manner similar to those of hnRNP A1, we next asked if H and A1 binding sites could cooperate to modulate 5' splice site selection. To answer this question, we produced a battery of pre-mRNA substrates containing combinations of hnRNP A1 and hnRNP H sites. Relative to individual sites, the combinations nearly always had the strongest impact on 5' splice site selection. Moreover, in at least two cases this impact was superior to the sum of the impact of individual binding sites, indicating that combinations of hnRNP H and hnRNP A1 binding sites can have a synergistic effect on 5' splice selection. However, some combinations of binding sites (i.e., HBS1-ABS1 and ABS1-HBS3) were less active than the reverse configurations and other combinations. The reason for these differences remains unclear and will require further investigations.

According to the model, the capacity of hnRNP A1 and hnRNP H binding sites to cooperate in 5' splice site selection implies that the hnRNP A1 and hnRNP H proteins can interact with one another. Although at least a fraction of hnRNP A1 and H proteins exist in the same complexes (Jeronimo et al. 2007), their ability to establish direct interactions has never been documented. We used the BRET technology to answer this question and our results clearly show that hnRNP A1 and hnRNP H proteins can co-exist in very close proximity, suggesting self-interaction and interaction with one another in live mammalian

cells. The C-terminal glycine rich domain (GRD) of A1 and H is important for both homotypic and heterotypic interactions since their removal strongly reduced the BRET signals. Moreover, the GRD of A1 and H are also essential for splicing activity and replacing the GRD of H with that of A1 maintained and even reinforced splicing activity.

Recently, we have provided further support for the looping out model by showing that intronic binding sites for hnRNP A1 or H can stimulate the splicing of longer introns *in vitro* (Martinez-Contreras et al. 2006). Since then, the looping model has been invoked to explain the activity of hnRNP A1 and other RNA binding proteins in splice site selection for other pre-mRNAs. For example, hnRNP A1 bound to exonic and intronic sites may cooperate to induce exon exclusion in the SMN2 pre-mRNA (Kashima et al. 2007). Also, the tissue-specific regulator Nova may multimerize to form an RNA loop to bring in closer proximity a distant 5' splice site and a branch site (Ule et al. 2006).

The capacity of hnRNP A1 and hnRNP H to interact and cooperate to modulate splicing decisions has important implications on our view of the mechanisms of splicing control. Given the abundance of these hnRNP proteins, splicing control through the combinatorial use of hnRNP A1 and hnRNP H binding sites may be a prevalent occurrence. In a recent study that has analyzed the impact of knocking down the major hnRNP proteins on the alternative splicing of 56 alternative splicing events, one cassette exon in CASP1 responded similarly to the depletion of hnRNP A2 and hnRNP H in HeLa cells (Venables et al. 2008). In an on-going larger screen in progress, we estimate that 5-10% of the alternative splicing events that react to an hnRNP A1 depletion are affected similarly by knocking down hnRNP H (Venables et al., unpublished data). This

partnership between A1 and H may also be used to stimulate the splicing of a large intron when the interacting proteins all bind to this intron.

Our finding also raises the possibility that hnRNP A1 and/or H can also entertain interactions with other hnRNP proteins, particularly those carrying glycine-rich domains (hnRNP F, G, P2, Q, D, U and R) (Martinez-Contreras et al. 2007). The binding of these proteins at various locations on a pre-mRNA or their association to other RNA binding proteins would expand the ways the conformation of a pre-mRNA could be rearranged to modulate splicing decisions.

MATERIALS AND METHODS

Plasmid constructs

Pre-mRNA substrates 553(p45), 553A1 -- (p42), 553-- A1 (p39) and 553A1 A1 (p36) have been described earlier (Nasim et al. 2002). To construct plasmids p553H1 -- and p553H1 A1, DNA oligonucleotide J1 (5'-CCATGGATCACTGGGGTGGATCATCCAGGTGGGGCTTTT-3') and J2 (5'-AAAAGCCCCACCTGGATGATCCACCCCAGTGATCCATGG-3') were hybridized and inserted into the SmaI sites of p45 and p39, and the EcoRV sites of p45 and p42 to generate p553--H1 and p553A1 H1, respectively. The same oligos were also inserted into the SmaI site of plasmids p553--H1 to generate p553H1 H1. Oligonucleotides J3 (5'-CGAATCGACAAAGGGGAGGAAGTGGGAGAAA-3') and J4 (5'-TTTCTCCCACTTCCTCCCCTTTGTCTGATTCG-3') were hybridized and cloned into the SmaI sites of p45 and p39 or introduced into the EcoRV sites of the p45 and p42 to produce p553H2--, p553H2 A1, p553--H2 and p553A1 H2, respectively. Plasmid p553--H2 was used as a backbone to introduce the complementary oligos into the SmaI site to generate p553H2 H2. Complementary oligos J5 (5'-CCATGGTTTGGGGGCAGTAGTTGG-3') and J6 (5'-CCAACTACTGCCCCCAAACCATGG-3') were inserted into the SmaI sites of the p45 and p39 or inserted into the EcoRV site of p45 and p42 to produce p553H3--, p553H3 A1, p553--H3 and p553A1H3, respectively. The same oligos were introduced into the SmaI site of p553--H3 to generate p553H3 H3.

Pre-mRNA substrates p553Hmut1 Hmut2, p553Hmut1 Hmut1 and pHmut2 Hmu1 were constructed using the same strategies presented earlier. Oligos (5'-CCATGGATCACTGGTGTGGATCATCCAGGTGGTGCTTTT-3') and (5'-CCATGGATCACTGGTGTGGATCATCCAGGTGGGGCTTTT -3') were hybridized with their complementary molecules and inserted into the SmaI sites of the p45 to generate the intermediate constructs p553Hmut1-- and pHmut2--. These inserts were then cloned into EcoRV of p553Hmut1-- and pHmut2-- to produce p553Hmut1 Hmut2, p553Hmut1 Hmut1 and p553Hmut2 Hmut1.

Mutant cDNAs Hs and Fs were produced by PCR amplification on pVL1392 hnRNP H and pVL1392 hnRNP F respectively. Forward primer F/H.1 (5'-CCGCTCCAGAATTCTAGAAATAATTTTGT-3') containing XbaI restriction site was used for both constructions. Reverse primer Hstop (5'-TTGGATCCCTACCCAGGTCTGTCA-3') and Fstop (5'-TTGGATCCCTACCCGGGCGG3') containing a BamHI site were used for Hs and Fs respectively. PCR fragments were cut with BamHI and XbaI and inserted in pVL1392 using the same restriction sites. Hybrid constructions H(qRRM1-2) + A1(GRD) (HA1) were generated by production of two intermediate plasmids. First, a SmaI restriction site was created in plasmid pVL1392 hnRNP H by PCR overlap mutagenesis. A first amplification from pVL1392 hnRNP H by primers F/H.1 and H2smaI (5'-CCCGGGTCTGTCATAAGG-3') and a second amplification from the same template by primers H3SmaI (5'-ACCCGGGGCTGGTAG-3') and H4 (5'-AAGGATCCCTATGCAATGTTTGATT-3') were performed. Products were purified and mixed for another PCR reaction using primers F/H.1 and H4. PCR fragment was

digested with XbaI and BamHI, cloned in pVL1392 to generate pVL1392 HSmaI. Another intermediate plasmid was created by insertion of an EagI restriction site in pQE80L-A1. PCR was performed from plasmid pQE80L-A1 using forward primer A1.1 (5'-ACCATCACGGATCCCCCATGTCTAAGTTCC-3') and reverse primer A1.2 EagI (5'-ACTGCGGCCGCGCTGACTGGATGAAGC-3'). Another reaction was performed with primer A1.3 EagI (5'-AGTCAGCGCGGCCGCGCAGTGGTTCTGG-3') and primer A1.4 (5'-AGCTAATTAAGCTTTGTTTCCTGGCTGTA-3'). The two PCR fragments were purified and mixed for another PCR using primer A1.1 and A1.4. The PCR amplicon was digested with HindIII and BamHI, cloned in pQE80L to generate pQE80L-A1 EagI. Fragment HindIII-EagI from this plasmid was removed and treated with Klenow. Plasmid pVL1392 HSmaI was digested with SmaI and BamHI and treated with Klenow. The SmaI-BamHI fragment was replaced with the HindIII-EagI fragment from pQE80L-A1 EagI to generate pVL1392 H-A1.

Plasmid pYFP H expressing YFP in C-terminal of hnRNP H was made by PCR amplification using pVL1392 hnRNP H and primers BRET H.1 (5'-TATCCGGGTACCGAGATGATGTTGGGCAC-3') and BRET YFP H.4 (5'-CGGATAGGATCCGCAATGTTTGATTGAAAATCAC-3'). PCR fragments were purified, cut with KpnI- BamHI and inserted in pYFP-CMV-Topaz using the same restriction sites. The same strategy was applied to create plasmids YFP Hs using primers BRET H.1 and BRET YFP H.5 (5'-CGGATAGGATCCGCCCCAGGTCTG-3'). Amplification on pQE80L-A1 with primers BRET A1.1 (5'-TATCCGGGTACCCCCATGTCTAAGTCCGAG-3') and BRET YFP A1.4 (5'-CGGATAGGATCCAACCTCCTGCCACTG-3') was used to generate plasmids YFP

A1, and amplification with primers BRET A1.1 and BRET YFP A1.5 (5'-CGGATAGGATCCCTGCGACCTCTCTGAC-3') to create plasmids YFP UP1. The PCR fragments were inserted in pYFP-CMV-Topaz using the same strategy. We performed a PCR reaction on pVL1392 A1H with primers BRET A1.1 and BRET YFP A1.4, a digestion to produce fragments that were inserted in pYFP-CMV-Topaz with the same restriction enzymes to produce YFP A1H. PCR amplification on pVL1392 H was carried out to create a plasmid that could express *Rluc* at the C-terminal portion of hnRNP H. Primer BRET H.1 and BRET LUC H.2 (5'-CGGATAGGATCCATGCAATGTTTGATTGAAAATCAC-3') were used. The amplified fragments were then purified, cut with KpnI-BamHI and inserted in pRluc-N1(h) using the same restriction sites. PCR reactions to generate *Rluc* A1 used oligos BRET A1.1 and BRET LUC A1.2 (5'-CGGATAGGATCCAGAACCTCCTGCCACTG-3') on pQE80L-A1. The PCR fragments resulting from this reaction were cut and inserted as described previously.

***In vitro* transcription and splicing assays**

Linearized constructs using ScaI were used as template for *in vitro* transcription. Pre-mRNA substrates were synthesized and purified as described (Martinez-Contreras et al. 2006). Two fmoles of substrates were incubated in HeLa nuclear extract (Dignam et al. 1983) under standard conditions (Chabot et al. 1997) at 30 °C. Purified RNA were amplified using Ready-To-Go RT-PCR beads (GE Healthcare) as described previously (Nasim et al. 2002). Nuclear extract prepared from HeLa cells knocked down for hnRNP F/H expression were produced using siRNA against hnRNP F/H (Garneau et al. 2005).

Recombinant His-tagged proteins were produced using baculovirus expression system (BD Biosciences) (Garneau et al. 2005).

Gel shift assay

RNA oligonucleotides (HBS1) were labeled using T4 polynucleotide kinase. RNA molecules were incubated in splicing mix (1.4 mM rATP, 9 mM MgCl₂, 50 mM creatine phosphate, 7% polyvinylalcohol, 5.5 mM dithiothreitol and 2 units of RNAGuard (GE Healthcare)). Recombinant proteins were then added to the mixture and allowed to incubate on ice for 5 min. Heparin at a final concentration of 0.75 mg/ml and loading dye were added and the samples were loaded on a 5% native acrylamide gel (29:1, acrylamide:bisacrylamide) in Tris-glycine buffer.

Filter binding assay

RNA oligonucleotides (HBS2, HBS3 and CE1a) were labeled using T4 polynucleotide kinase. After the reaction was completed, quantitated and the RNA purified, cold RNA oligo was added to obtain equivalent specific activity (cpm/pmole) for the different labeled RNA oligos. We then incubated these labeled RNAs with the recombinant proteins as previously described for the gel shift assay. Mixtures were then added to a 96-well filtration apparatus (Bio-Rad Bio-Dot) and washed with buffer containing 10 mM Hepes pH 7.9, 50 mM KCl, 10% glycerol and 0.1 mM EDTA.

BRET assay

EcR 293 cells were seeded in 12-well plates at a density of 1.5×10^5 cells/well. Twenty-four hours later transfection was accomplished using lipofectamine 2000

(Invitrogen). Ten ng of plasmids *Rluc*, *Rluc H* or *Rluc A1* were co-transfected with increasing amounts (50, 100, 250 and 500 ng) of plasmids YFP, YFP H, YFP Hs, YFP A1, YFP UP1 or YFP A1H. Thirty-six hours post-transfection, cells were collected, washed and resuspended in 200 μ l of PBS 1X. Ninety μ l were distributed in a 96-well microplate. Expression level of the YFP fused proteins were determined by measurement of fluorescence (YFP) (excitation filter at 485 nm and detection filter at 550 nm). The BRET signal is calculated as the fluorescence/luminescence after addition of 5 μ M coelenterazine h, which is a substrate for luciferase (emission filter 1 for YFP at 550 nm and emission filter 2 for *Rluc* at 470 nm). Measurements were realized with the Fusion alpha-FP apparatus (PerkinElmer-Canberra, Packard Bio Science).

ACKNOWLEDGMENTS

This study was supported by a grant from the Canadian Institute of Health Research to B.C. B.C. is a Canada Research Chair in Functional Genomics. S.D.-B. was supported by a studentship from the Fonds pour la recherche en santé du Québec (FRSQ).

REFERENCES

- Arhin, G.K., Boots, M., Bagga, P.S., Milcarek, C., and Wilusz, J. 2002. Downstream sequence elements with different affinities for the hnRNP H/H' protein influence the processing efficiency of mammalian polyadenylation signals. *Nucleic Acids Res.* 30: 1842-1850.
- Blanchette, M. and Chabot, B. 1999. Modulation of exon skipping by high-affinity hnRNP A1-binding sites and by intron elements that repress splice site utilization. *EMBO J.* 18: 1939-1952.
- Buratti, E., Baralle, M., De Conti, L., Baralle, D., Romano, M., Ayala, Y.M., and Baralle, F.E. 2004. hnRNP H binding at the 5' splice site correlates with the pathological effect of two intronic mutations in the NF-1 and TSHbeta genes. *Nucleic Acids Res.* 32: 4224-4236.
- Burd, C.G. and Dreyfuss, G. 1994. RNA binding specificity of hnRNP A1: significance of hnRNP A1 high-affinity binding sites in pre-mRNA splicing. *EMBO J.* 13: 1197-1204.
- Caputi, M. and Zahler, A.M. 2001. Determination of the RNA binding specificity of the heterogeneous nuclear ribonucleoprotein (hnRNP) H/H'/F/2H9 family. *J. Biol. Chem.* 276: 43850-43859.
- Caputi, M. and Zahler, A.M. 2002. SR proteins and hnRNP H regulate the splicing of the HIV-1 tev-specific exon 6D. *EMBO J.* 21: 845-855.

- Cartegni, L., Maconi, M., Morandi, E., Cobianchi, F., Riva, S., and Biamonti, G. 1996. hnRNP A1 selectively interacts through its Gly-rich domain with different RNA-binding proteins. *J. Mol. Biol.* 259: 337-348.
- Chabot, B., Blanchette, M., Lapierre, I., and La Branche, H. 1997. An intron element modulating 5' splice site selection in the hnRNP A1 pre-mRNA interacts with hnRNP A1. *Mol. Cell Biol.* 17: 1776-1786.
- Chabot, B., LeBel, C., Hutchison, S., Nasim, F.H., and Simard, M.J. 2003. Heterogeneous nuclear ribonucleoprotein particle A/B proteins and the control of alternative splicing of the mammalian heterogeneous nuclear ribonucleoprotein particle A1 pre-mRNA. *Prog. Mol. Subcell. Biol.* 31: 59-88.
- Chou, M.Y., Rooke, N., Turck, C.W., and Black, D.L. 1999. hnRNP H is a component of a splicing enhancer complex that activates a c-src alternative exon in neuronal cells. *Mol. Cell Biol.* 19: 69-77.
- Crawford, J.B. and Patton, J.G. 2006. Activation of alpha-tropomyosin exon 2 is regulated by the SR protein 9G8 and heterogeneous nuclear ribonucleoproteins H and F. *Mol. Cell Biol.* 26: 8791-8802.
- Dallaire, F., Dupuis, S., Fiset, S., and Chabot, B. 2000. Heterogeneous nuclear ribonucleoprotein A1 and UP1 protect mammalian telomeric repeats and modulate telomere replication in vitro. *J. Biol. Chem.* 275: 14509-14516.

Dignam, J.D., Lebovitz, R.M., and Roeder, R.G. 1983. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.* 11: 1475-1489.

Engelbrecht, J., Knudsen, S., and Brunak, S. 1992. G+C-rich tract in 5' end of human introns. *J. Mol. Biol.* 227: 108-113.

Expert-Bezancon, A., Sureau, A., Durosay, P., Salesse, R., Groeneveld, H., Lecaer, J.P., and Marie, J. 2004. hnRNP A1 and the SR proteins ASF/SF2 and SC35 have antagonistic functions in splicing of beta-tropomyosin exon 6B. *J. Biol. Chem.* 279: 38249-38259.

Fogel, B.L. and McNally, M.T. 2000. A cellular protein, hnRNP H, binds to the negative regulator of splicing element from Rous sarcoma virus. *J. Biol. Chem.* 275: 32371-32378.

Garneau, D., Revil, T., Fisette, J.F., and Chabot, B. 2005. Heterogeneous nuclear ribonucleoprotein F/H proteins modulate the alternative splicing of the apoptotic mediator Bcl-x. *J. Biol. Chem.* 280: 22641-22650.

Graveley, B.R. 2000. Sorting out the complexity of SR protein functions. *RNA* 6: 1197-1211.

Graveley, B.R. 2001. Alternative splicing: increasing diversity in the proteomic world. *Trends Genet.* 17: 100-107.

Graveley, B.R., Hertel, K.J., and Maniatis, T. 2001. The role of U2AF35 and U2AF65 in enhancer-dependent splicing. *RNA* 7: 806-818.

Guil, S., Gattoni, R., Carrascal, M., Abian, J., Stevenin, J., and Bach-Elias, M. 2003.

Roles of hnRNP A1, SR proteins, and p68 helicase in c-H-ras alternative splicing regulation. *Mol. Cell. Biol.* 23: 2927-2941.

Han, K., Yeo, G., An, P., Burge, C.B., and Grabowski, P.J. 2005. A combinatorial code for splicing silencing: UAGG and GGGG motifs. *PLoS Biol.* 3: e158.

Honore, B., Rasmussen, H.H., Vorum, H., Dejgaard, K., Liu, X., Gromov, P., Madsen, P., Gesser, B., Tommerup, N., and Celis, J.E. 1995. Heterogeneous nuclear ribonucleoproteins H, H', and F are members of a ubiquitously expressed subfamily of related but distinct proteins encoded by genes mapping to different chromosomes. *J. Biol. Chem.* 270: 28780-28789.

Hutchison, S., LeBel, C., Blanchette, M., and Chabot, B. 2002. Distinct sets of adjacent heterogeneous nuclear ribonucleoprotein (hnRNP) A1/A2 binding sites control 5' splice site selection in the hnRNP A1 mRNA precursor. *J. Biol. Chem.* 277: 29745-29752.

Jacquet, S., Mereau, A., Bilodeau, P.S., Damier, L., Stoltzfus, C.M., and Branlant, C. 2001. A second exon splicing silencer within human immunodeficiency virus type 1 tat exon 2 represses splicing of Tat mRNA and binds protein hnRNP H. *J. Biol. Chem.* 276: 40464-40475.

Jeronimo, C., Forget, D., Bouchard, A., Li, Q., Chua, G., Poitras, C., Therien, C., Bergeron, D., Bourassa, S., Greenblatt, J. et al. 2007. Systematic analysis of the protein interaction network for the human transcription machinery reveals the identity of the 7SK capping enzyme. *Mol. Cell.* 27: 262-274.

- Johnson, J.M., Castle, J., Garrett-Engle, P., Kan, Z., Loerch, P.M., Armour, C.D., Santos, R., Schadt, E.E., Stoughton, R., and Shoemaker, D.D. 2003. Genome-wide survey of human alternative pre-mRNA splicing with exon junction microarrays. *Science* 302: 2141-2144.
- Kanopka, A., Muhlemann, O., and Akusjarvi, G. 1996. Inhibition by SR proteins of splicing of a regulated adenovirus pre-mRNA. *Nature* 381: 535-538.
- Kashima, T. and Manley, J.L. 2003. A negative element in SMN2 exon 7 inhibits splicing in spinal muscular atrophy. *Nat. Genet.* 34: 460-463.
- Kashima, T., Rao, N., and Manley, J.L. 2007. An intronic element contributes to splicing repression in spinal muscular atrophy. *Proc. Natl. Acad. Sci.* 104: 3426-3431.
- Kohtz, J.D., Jamison, S.F., Will, C.L., Zuo, P., Luhrmann, R., Garcia-Blanco, M.A., and Manley, J.L. 1994. Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. *Nature* 368: 119-124.
- Lavigne, A., La Branche, H., Kornblihtt, A.R., and Chabot, B. 1993. A splicing enhancer in the human fibronectin alternate ED1 exon interacts with SR proteins and stimulates U2 snRNP binding. *Genes Dev.* 7: 2405-2417.
- Majewski, J. and Ott, J. 2002. Distribution and characterization of regulatory elements in the human genome. *Genome Res.* 12: 1827-1836.
- Martinez-Contreras, R., Cloutier, P., Shkreta, L., Fiset, J.F., Revil, T., and Chabot, B. 2007. hnRNP proteins and splicing control. *Adv. Exp. Med. Biol.* 623: 123-147.

- Martinez-Contreras, R., Fisette, J.F., Nasim, F.U., Madden, R., Cordeau, M., and Chabot, B. 2006. Intronic binding sites for hnRNP A/B and hnRNP F/H proteins stimulate pre-mRNA splicing. *PLoS Biol.* 4: e21.
- Mauger, D.M., Lin, C., and Garcia-Blanco, M.A. 2008. hnRNP H and hnRNP F complex with Fox2 to silence fibroblast growth factor receptor 2 Exon IIIc. *Mol. Cell. Biol.* 28: 5403-5419.
- Mayeda, A., Munroe, S.H., Caceres, J.F., and Krainer, A.R. 1994. Function of conserved domains of hnRNP A1 and other hnRNP A/B proteins. *EMBO J.* 13: 5483-5495.
- McCullough, A.J. and Berget, S.M. 1997. G triplets located throughout a class of small vertebrate introns enforce intron borders and regulate splice site selection. *Mol. Cell Biol.* 17: 4562-4571.
- Min, H., Chan, R.C., and Black, D.L. 1995. The generally expressed hnRNP F is involved in a neural-specific pre-mRNA splicing event. *Genes Dev.* 9: 2659-2671.
- Nasim, F.U., Hutchison, S., Cordeau, M., and Chabot, B. 2002. High-affinity hnRNP A1 binding sites and duplex-forming inverted repeats have similar effects on 5' splice site selection in support of a common looping out and repression mechanism. *RNA* 8: 1078-1089.
- Nussinov, R. 1987. (A)GGG(A), (A)CCC(A) and other potential 3' splice signals in primate nuclear pre-mRNA sequences. *Biochim. Biophys. Acta* 910: 261-270.
- Nussinov, R. 1988. Conserved quartets near 5' intron junctions in primate nuclear pre-mRNA. *J. Theor. Biol.* 133: 73-84.

- Nussinov, R. 1989. Conserved signals around the 5' splice sites in eukaryotic nuclear precursor mRNAs: G-runs are frequent in the introns and C in the exons near both 5' and 3' splice sites. *J. Biomol. Struct. Dyn.* 6: 985-1000.
- Paradis, C., Cloutier, P., Shkreta, L., Toutant, J., Klarskov, K., and Chabot, B. 2007. hnRNP I/PTB can antagonize the splicing repressor activity of SRp30c. *RNA* 13: 1287-1300.
- Pfleger, K.D., Seeber, R.M., and Eidne, K.A. 2006. Bioluminescence resonance energy transfer (BRET) for the real-time detection of protein-protein interactions. *Nat. Protoc.* 1: 337-345.
- Romano, M., Marcucci, R., Buratti, E., Ayala, Y.M., Sebastio, G., and Baralle, F.E. 2002. Regulation of 3' splice site selection in the 844ins68 polymorphism of the cystathionine Beta -synthase gene. *J. Biol. Chem.* 277: 43821-43829.
- Rooke, N., Markovtsov, V., Cagavi, E., and Black, D.L. 2003. Roles for SR proteins and hnRNP A1 in the regulation of c-src exon N1. *Mol. Cell. Biol.* 23: 1874-1884.
- Simard, M.J. and Chabot, B. 2002. SRp30c is a repressor of 3' splice site utilization. *Mol. Cell. Biol.* 22: 4001-4010.
- Solovyev, V.V., Salamov, A.A., and Lawrence, C.B. 1994. Predicting internal exons by oligonucleotide composition and discriminant analysis of spliceable open reading frames. *Nucleic Acids Res.* 22: 5156-5163.

- Ule, J., Stefani, G., Mele, A., Ruggiu, M., Wang, X., Taneri, B., Gaasterland, T., Blencowe, B.J., and Darnell, R.B. 2006. An RNA map predicting Nova-dependent splicing regulation. *Nature* 444: 580-586.
- Venables, J.P., Koh, C.S., Froehlich, U., Lapointe, E., Couture, S., Inkel, L., Bramard, A., Paquet, E.R., Watier, V., Durand, M. et al. 2008. Multiple and specific mRNA processing targets for the major human hnRNP proteins. *Mol. Cell Biol.* 28: 6033-6043.
- Wang, Z., Hoffmann, H.M., and Grabowski, P.J. 1995. Intrinsic U2AF binding is modulated by exon enhancer signals in parallel with changes in splicing activity. *RNA* 1: 21-35.
- Xu, Y., Piston, D.W., and Johnson, C.H. 1999. A bioluminescence resonance energy transfer (BRET) system: application to interacting circadian clock proteins. *Proc. Natl. Acad. Sci.* 96: 151-156.
- Yeo, G., Hoon, S., Venkatesh, B., and Burge, C.B. 2004. Variation in sequence and organization of splicing regulatory elements in vertebrate genes. *Proc. Natl. Acad. Sci.* 101: 15700-15705.
- Zahler, A.M., Damgaard, C.K., Kjems, J., and Caputi, M. 2004. SC35 and heterogeneous nuclear ribonucleoprotein A/B proteins bind to a juxtaposed exonic splicing enhancer/exonic splicing silencer element to regulate HIV-1 tat exon 2 splicing. *J. Biol. Chem.* 279: 10077-10084.

Zuo, P. and Maniatis, T. 1996. The splicing factor U2AF35 mediates critical protein-protein interactions in constitutive and enhancer-dependent splicing. *Genes Dev.* 10: 1356-1368.

DISCUSSION

L'objectif principal de notre projet consistait à étudier le contrôle de l'épissage alternatif par la protéine hnRNP H. Nos travaux étaient basés sur les démonstrations initiales qui ont établi que le facteur hnRNP A1 est impliqué dans l'exclusion de l'exon alternatif 7B de son propre ARN pré-messager. Plus précisément, le modèle pouvant expliquer cette modulation consiste en la formation d'homodimères entre molécules hnRNP A1, liées de chaque côté de l'exon 7B, qui permet le rapprochement des sites d'épissage 5' et 3' de l'exon 7 et 8 respectivement. Les expériences que nous avons effectuées suggèrent que des interactions homotypiques entre molécules hnRNP H et hétérotypiques entre les protéines hnRNP A1 et hnRNP H peuvent modifier l'architecture de l'ARN pré-messager afin de moduler la sélection des sites d'épissage 5'.

L'activité des sites de liaison pour hnRNP H

Les expériences que nous avons effectuées ont permis d'établir un modèle de contrôle de l'épissage alternatif impliquant hnRNP H et dont le mécanisme est calqué sur celui de hnRNP A1. Nos travaux qui supportent une telle affirmation ont été réalisés en utilisant une unité d'épissage artificielle constituée de deux sites d'épissage 5' en compétition avec un seul site d'épissage 3'. Lorsque des motifs de liaison pour hnRNP H sont introduits dans cette unité d'épissage, la sélection des sites d'épissage 5' est déplacée vers le distal. Un aspect qui a été tenu en considération dans cette première observation est la possibilité que l'activité observée ne soit pas causée spécifiquement par les éléments introduits. Par exemple, l'introduction d'un motif à un endroit précis de l'ARN pré-messager pourrait détruire une séquence régulatrice déjà présente ou en créer une autre. De plus, cette insertion pourrait rapprocher ou éloigner un motif de régulation d'un site

d'épissage, modifiant ainsi l'effet positif ou négatif de l'élément sur le recrutement de la machinerie du « spliceosome ». Les résultats que nous avons obtenus appuient l'hypothèse d'une participation directe de nos éléments liés par hnRNP H dans la modulation des sites d'épissage 5'. Tout d'abord, des unités d'épissage contenant des mutations dans le quadruplet GGGG engendrent une augmentation significative du pourcentage de produits proximaux par rapport à l'ARN pré-messager contenant la séquence de type sauvage (Figure 1B, Chapitre 2). La seconde approche que nous avons utilisée postule que nous pouvons reproduire l'activation du site d'épissage 5' distal en utilisant des séquences contenant d'autres dispositions d'éléments riches en guanosines (capables de recruter hnRNP H), et dont les nucléotides adjacents sont différents (Figure 2A, Chapitre 2). L'insertion de ces séquences dans notre ARN pré-messager augmente effectivement le pourcentage de produits distaux (Figure 2C et 2D, Chapitre 2), confirmant à nouveau l'importance des éléments riches en guanosines dans la sélection des sites d'épissage. Ces observations sont en accord avec les travaux précédents qui ont déjà répertorié l'importance des motifs riches en guanosines dans la modulation des sites d'épissage en compétition. Par exemple, une unité d'épissage provenant du gène de l' *α -globine*, contenant deux sites d'épissage 5' en compétition, utilise exclusivement le site distal lorsque des éléments GGG sont introduits de par et d'autre du site proximal (McCullough et Berget, 1997). Dans une autre situation, des séquences riches en guanosines peuvent aussi moduler la sélection de sites d'épissage 3' en compétition dans le gène de la *cystathionine β -synthase* (Romano et al., 2002).

Notre hypothèse initiale dans la conception de notre étude impliquait que les molécules hnRNP H soient recrutées aux sites de liaison introduits et activent l'utilisation

du site d'épissage 5' distal. Nous nous devions d'écarter la possibilité que la formation d'une structure secondaire, suite à l'insertion de nos séquences, soit en cause dans l'activité d'épissage observée. Les motifs de liaison pour la protéine hnRNP H correspondent à des séquences riches en guanosines qui sont de parfaits candidats pour la formation d'un G-quadruplex. Cette structure est causée par la formation de liens hydrogènes de type « Hoogsteen » entre des guanosines qui sont disposés en répétition (figure 9)(Keniry, 2000). Un G-quadruplex peut permettre la liaison de protéines et être impliqué dans la modulation de l'épissage alternatif. Par exemple, chez le gène *FMR* (« fragile X mental retardation »), la liaison du facteur FMRP (« fragile X mental retardation protein ») à cette structure serait impliquée dans l'activation d'un exon alternatif (Didiot et al., 2008). Dans le cadre de notre projet, une telle structure aurait pu influencer la sélection des sites d'épissage et ce, à plusieurs niveaux. Par exemple, un G-

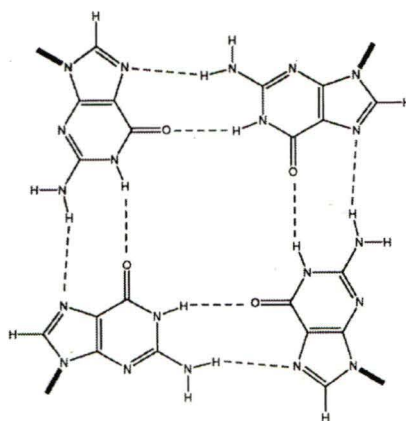


Figure 9. Représentation d'une structure G-quadruplex. Une séquence contenant des répétitions de guanosines peut engendrer la formation de G-quadruplex par l'intermédiaire de liens hydrogènes de type « Hoogsteen » entre ces nucléotides. Tiré de (Keniry, 2000).

quadruplex pourrait recruter une protéine autre que hnRNP H et influencer la sélection des sites d'épissage 5'. Une seconde possibilité serait que la structure déstabilise la liaison

de facteurs à proximité ou inhibe le recrutement de la machinerie d'épissage. Les expériences que nous avons effectuées nous permettent de confirmer que la protéine hnRNP H est directement impliquée dans la modulation des sites d'épissage 5'. En effet, l'activité des sites de liaison pour hnRNP H est compromise lorsque l'ARN pré-messager est incubé dans un extrait nucléaire dont la concentration en protéines hnRNP H a été diminuée (figure 6B, Martinez-Contreras *et al.*, 2006). De plus, lorsqu'une quantité croissante de protéines recombinantes hnRNP H est ajoutée à cet extrait, nous détectons une augmentation du ratio distal/proximal (figure 6B, Martinez-Contreras *et al.*, 2006). Ces expériences appuient ainsi la notion d'une participation directe de ce facteur dans l'activation du site d'épissage 5' distal. Un aspect que nous ne pouvons pas écarter est la possibilité que le recrutement de la protéine hnRNP H s'effectue via un G-quadruplex. Jusqu'à présent, aucune évidence n'a démontré la nécessité d'une telle structure pour la liaison de ce facteur. Cependant, cette avenue mériterait d'être approfondie et pourrait permettre de mieux comprendre comment le régulateur hnRNP H peut être recruté à l'ARN.

Un argument majeur en faveur du modèle que nous voulons démontrer est la capacité des sites de liaison pour hnRNP H de coopérer pour activer la sélection du site d'épissage 5' distal. Nos résultats nous montrent que l'augmentation de l'utilisation de ce site est maximale lorsque les motifs de liaison pour hnRNP H sont introduits en combinaison dans notre unité d'épissage (figure 5C, Martinez-Contreras *et al.*, 2006)(figure 2C,D, Chapitre 2). Lorsque nous additionnons l'activité des unités d'épissage qui contiennent un seul élément (en aval et en amont du site d'épissage 5' proximal), nous constatons que cette valeur est toujours inférieure à celle produite par une

combinaison de sites sur un seul ARN pré-messager (figure 2E, 2F et 2G, Chapitre 2).

Cette observation impliquerait que ces séquences liées par hnRNP H coopèreraient afin de déplacer en amont le choix du site d'épissage. Des résultats réalisés par d'autres groupes sont aussi compatibles avec un tel mécanisme. Par exemple, l'exclusion d'un exon du gène *SMN2* (« survival motor neuron ») serait dictée par deux éléments (l'un exonique et l'autre intronique) qui recruteraient hnRNP A1 et agiraient en coopération (Kashima et al., 2007). Une autre étude a montré que les séquences UAGG et GGGG peuvent coopérer pour inactiver l'inclusion d'exons alternatifs (Han et al., 2005).

Ce comportement de nos unités d'épissage est en accord avec le modèle que notre laboratoire a proposé pour hnRNP A1 qui implique un rapprochement de sites d'épissage distants par l'intermédiaire d'interactions entre molécules liées à l'ARN. Ce mécanisme implique que l'insertion d'un seul site de liaison pour hnRNP A1, en amont ou en aval d'un site d'épissage 5', ne devrait pas modifier les quantités relatives d'isoformes. Cependant, dans la majorité des situations, nous observons une activité en présence d'un seul élément : c'est le cas, par exemple, des constructions 553H2-- et 553--H2 dont le pourcentage de produits proximaux diminue comparativement au profil d'épissage de l'unité 553 (Figure 2C, piste 1, 2 et 3, Chapitre 2). Nous attribuons cette activité à la possibilité que la protéine hnRNP H, liée au site fort HBS, puisse interagir avec d'autres molécules hnRNP H ou hnRNP A1 liées à des sites plus faibles dans notre ARN pré-messager. En appui à cette hypothèse, nous constatons que l'unité d'épissage que nous avons utilisée comprend plusieurs séquences riches en guanosines, offrant ainsi des sites de liaison potentiels pour ces protéines. De plus, il a été montré que le facteur hnRNP A1 peut interagir avec certains sites d'épissage 3', dont celui de l'adénovirus, présent dans

notre ARN pré-messager (Buvoli et al., 1990b). Il était impossible pour nous de muter les éléments requis pour la liaison de hnRNP A1 à cette séquence, puisqu'immanquablement nous allions détruire le site d'épissage. Par contre, nous envisageons de remplacer cette séquence d'adénovirus par d'autres sites d'épissage 3' tout en s'assurant qu'aucune protéine hnRNP A1 ne pourra lier.

Afin d'appuyer nos observations qui suggèrent que des combinaisons de sites de liaison pour hnRNP H peuvent coopérer, nous avons montré que la modulation de la sélection des sites d'épissage 5' ne s'effectue pas par une activation ou inhibition directe de l'assemblage de la snRNP U1. Des travaux antécédents ont montré que la présence de sites de liaison pour hnRNP A1, dans l'unité d'épissage 553, n'affecte pas le recrutement de la snRNP U1 (Chabot et al., 1997). Dans le cadre de notre étude, des expériences de protection à la RNase H révèlent une légère variation de la liaison de la snRNP U1 lorsque des sites de liaison pour hnRNP H sont introduits. Par contre, nous ne constatons aucune corrélation directe entre cette petite modification de la stabilité de la snRNP U1 et le fort changement dans la sélection des sites d'épissage 5' (figure 6D, Martinez-Contreras *et al.*, 2006). Cette observation nous suggère donc que l'activité des sites de liaison pour hnRNP H ne peut pas passer par une activation ou inactivation directe des sites d'épissage 5'. Il n'en demeure pas moins qu'un site d'épissage 5' proximal, bordé de motifs de liaison pour hnRNP H, est limité à s'engager dans les deux réactions de transestérification. Puisque notre modèle propose un rapprochement de séquences distantes, grâce à l'interaction de protéines liées à l'ARN, le site d'épissage 5' proximal compris entre ces deux régions jointes pourrait être masqué. Des études ont montré qu'une unité d'épissage contenant des motifs de liaison pour hnRNP A1, de par et d'autre d'un site d'épissage 5',

peut inactiver l'utilisation de celui-ci en compromettant le passage au complexe B (correspondant à l'arrivée de la tri-snRNP U4/U5/U6) (Nasim et al., 2002). Lorsque les motifs de liaison pour hnRNP A1 sont remplacés par des séquences complémentaires, l'appariement des bases nucléotidiques permet de simuler l'effet de changement de conformation de l'ARN pré-messager et de reproduire l'activité d'épissage (Nasim et al., 2002). De plus, cette activité est associée à l'inhibition de l'assemblage du complexe B (Nasim et al., 2002). Considérant ces observations, nous pourrions postuler que le mécanisme de répression du site d'épissage 5' proximal, des ARN pré-messagers constitués de combinaisons de sites de liaison pour hnRNP H, passerait par une inhibition de l'assemblage de la tri-snRNP U4/U5/U6.

Étant donné que des paires de sites de liaison pour les protéines hnRNP A1 ou hnRNP H peuvent coopérer et activer l'utilisation du site d'épissage 5' distal, nous avons vérifié si une combinaison hétérotypique de sites de liaison pour ces deux molécules pouvait avoir une activité. La figure 3B (Chapitre 2) nous montre que l'on peut stimuler l'utilisation du site d'épissage 5' distal avec plusieurs combinaisons de sites de liaison pour hnRNP H et hnRNP A1. De plus, étant donné que la somme des activités des ARN pré-messagers contenant un site est inférieure à celle obtenue par une combinaison de motifs, (figure 3C et 3D, piste 2 vs 3, Chapitre 2) une coopération entre les séquences liées par hnRNP H et hnRNP A1 expliquerait l'activité observée. Il est intéressant de constater que l'activation du site d'épissage 5' distal peut passer par différents arrangements de motifs liés par hnRNP A1 et hnRNP H. Cette variété de combinaisons peut engendrer des niveaux d'activité différents : nous pouvons observer que dans tous les cas, une combinaison de motifs de liaison pour hnRNP A1 est plus active qu'une

disposition hétérotypique (entre hnRNP H et hnRNP A1) ou homotypique (entre hnRNP H). Ces degrés de stimulation qui varient en fonction des régulateurs pourraient avoir des répercussions importantes dans l'exécution de processus biologiques. Par exemple, l'ajustement des niveaux relatifs de certains isoformes pourrait être contrôlé par le recrutement d'une combinaison précise de molécules hnRNP H et/ou hnRNP A1 qui, selon l'activité de celle-ci, permettrait de générer des quantités précises d'ARN messagers codant pour une protéine particulière.

Nos travaux effectués portant sur l'épissage des longs introns viennent appuyer le modèle de coopération entre molécules hnRNP H et hnRNP A1 (Martinez-Contreras *et al.*, 2006). L'utilisation d'une unité d'épissage contenant un intron artificiellement élargi nous a permis de montrer que l'insertion de sites de liaison pour la protéine hnRNP A1, aux extrémités de l'intron, peut stimuler l'épissage de cette ARN pré-messager (Martinez-Contreras *et al.*, 2006). De plus, des sites de liaison pour hnRNP H peuvent reproduire cette stimulation : celle-ci était observée seulement lorsque les séquences étaient introduites en combinaison, près des extrémités de l'intron, suggérant une coopération entre les sites de liaison (Martinez-Contreras *et al.*, 2006). Le modèle retenu pouvant expliquer cette stimulation implique le rapprochement des sites d'épissage 5' et 3' par l'intermédiaire d'interactions entre les protéines liées aux extrémités du grand intron (Martinez-Contreras *et al.*, 2006). Ce mécanisme proposé permettrait d'expliquer comment un complexe au site d'épissage 5' peut trouver ses partenaires au site d'épissage 3' tout en évitant la sélection de pseudos sites retrouvés dans l'intron (figure 10).

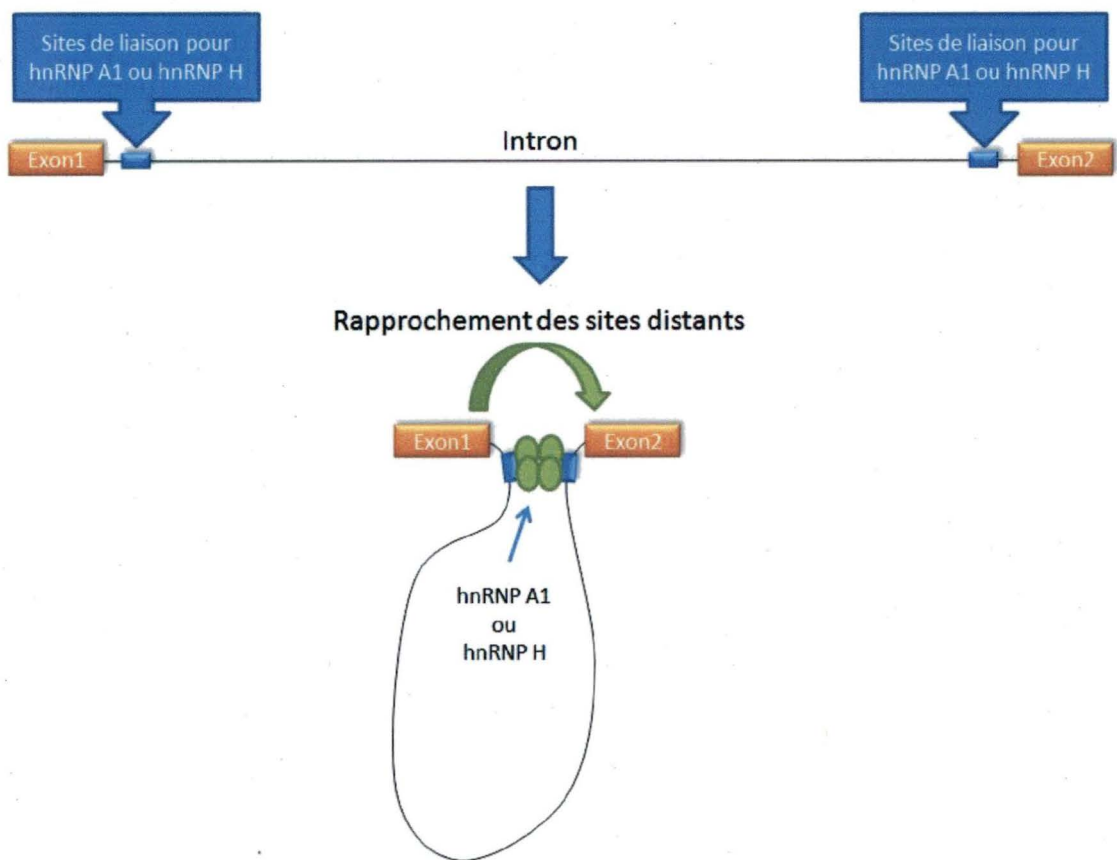


Figure 10. Stimulation de l'épissage d'un ARN pré-messager. Les protéines hnRNP A1 et hnRNP H peuvent activer l'épissage d'une unité contenant un long intron par interaction entre ces molécules liées aux extrémités de l'intron.

Interactions homotypiques et hétérotypiques

Notre modèle prédit que la coopération entre les sites de liaison pour hnRNP H et hnRNP A1 s'effectue par l'intermédiaire d'interactions entre les protéines liées à l'ARN qui rapprochent des sites d'épissage 5' et 3' distants. Des études ont documenté que des molécules hnRNP A1 peuvent interagir et que le domaine riche en résidus glycines est important pour la formation du complexe (Cartegni et al., 1996). De plus, des expériences d'immunoprécipitation et des analyses de spectrométrie de masse ont révélé que la molécule hnRNP H peut interagir avec hnRNP A1 (Jeronimo et al., 2007). Nous avons

utilisé la technologie BRET (« bioluminescence resonance energy transfert ») afin d'étudier les interactions homotypiques et hétérotypiques entre les protéines hnRNP H et hnRNP A1 dans les cellules vivantes. Ce type d'expérience est basé sur le transfert d'énergie entre une molécule étiquetée Luciférase et une autre YFP (« yellow fluorescent protein »). Le transfert d'énergie s'effectue lorsque les deux partenaires protéiques se situent dans un contexte de forte proximité (50 à 100 Å) correspondant à la distance typique entre deux protéines qui interagissent. Dans un premier temps, nous avons confirmé les interactions entre deux protéines hnRNP A1, de même qu'entre hnRNP H et hnRNP A1 (figure 5B-E, Chapitre 2). Nos expériences de BRET nous ont aussi démontré que deux protéines hnRNP H peuvent se retrouver très près l'une de l'autre dans une cellule et que le signal qui décrit cette proximité est dépendant de la région en position C-terminale de cette protéine (figure 5D et 5E, Chapitre 2). Ces expériences effectuées par l'intermédiaire de la technologie BRET constituent, à notre connaissance, la première évidence que des interactions homotypiques et hétérotypiques entre molécules hnRNP H et hnRNP A1 peuvent se former dans des cellules vivantes. Cette observation soulève un questionnement quant à la chronologie des événements menant à la liaison des molécules à l'ARN et la formation d'interactions entre celles-ci. Une étude structurale effectuée à l'aide de la protéine UP1 (version mutante de hnRNP A1 qui est dépourvue de son domaine riche en résidus glycines) pourrait nous aider à mieux comprendre ce mécanisme. Celle-ci révèle que la molécule UP1 peut interagir avec deux motifs situés sur deux oligos distincts et que le RRM 1 et RRM 2 seraient importants pour établir la liaison à la première et seconde séquence respectivement (Ding et al., 1999). Basé sur ces observations, il a été proposé que, suite à la liaison des molécules hnRNP A1 à l'ARN, il y aurait formation de partenaires d'interaction par l'intermédiaire des GRD et ce

complexe serait ensuite stabilisé via la participation des deux RRM qui lieraient les deux motifs d'ARN distincts (Chabot et al., 2003).

Au même titre que hnRNP A1, la région en position C-terminale de hnRNP H est aussi importante pour permettre des interactions homotypiques et hétérotypiques. En effet, la substitution de la section en position C-terminale de hnRNP A1, par celle de hnRNP H (qui résulte en la protéine hybride A1H), permet de restaurer le signal BRET suite à la délétion du GRD de hnRNP A1 (figure 5C et 5E, Chapitre 2). Une corrélation directe entre la formation d'interactions protéines : protéines et la capacité d'activer l'utilisation du site d'épissage 5' distal peut être établie. En effet, la section en position C-terminale (correspondant au GRD) de hnRNP A1 est à la fois importante pour lui permettre de former des homodimères et pour son activité dans l'épissage (Blanchette et Chabot, 1999). La même observation peut être faite avec hnRNP H quant à la région correspondant aux deux GRD et au qRRM3. Alors que les expériences de BRET nous révèlent qu'une protéine hnRNP H ayant une délétion de cette section (mutant Hs) ne peut pas interagir avec la version type sauvage de hnRNP H, l'ajout du mutant Hs à l'extrait nucléaire $\Delta F/H$ n'a aucun effet sur la sélection des sites d'épissage de l'unité 553H1H1 (figure 5D, piste 3 et 4 vs piste 9 et 10, Chapitre 2). Contrairement à hnRNP A1, hnRNP H contient un qRRM 3 qui entrecoupe les deux GRD. Les résultats que nous avons obtenus suggèrent fortement que les deux domaines riches en résidus glycines de hnRNP H jouent un rôle majeur dans l'activation du site d'épissage 5' distal. En effet, la substitution de la région contenant les deux GRD et le qRRM3, par le GRD de hnRNP A1, permet de maintenir et même augmenter l'activité d'épissage (figure 5D, piste 13 et 14, Chapitre 2). Néanmoins, nous n'excluons pas la possibilité que le qRRM3 de hnRNP

H soit important pour la formation d'interactions protéines : protéines et, conséquemment, être impliqué dans la sélection des sites d'épissage. Malgré le fait que les résidus d'acides aminés contenus dans ce domaine correspondent à un motif de liaison à l'ARN, la participation d'un RRM dans une interaction entre molécules a déjà été documentée. Par exemple, la protéine U2AF65 contient deux RRM impliqués dans la liaison à la séquence riche en pyrimidines au site d'épissage 3' et un troisième RRM, en position C-terminale, important pour son interaction avec la protéine SF1 au site de branchement (Berglund et al., 1998).

Génomique humaine et modulation de l'épissage

Le modèle dans lequel des protéines hnRNP peuvent interagir, suite à la liaison à l'ARN, pour moduler la sélection des sites d'épissage 5' est basé sur plusieurs observations. Tout d'abord, des éléments de liaison pour hnRNP H et hnRNP A1 peuvent coopérer pour stimuler l'utilisation du site d'épissage 5' distal sans déstabiliser de façon significative le recrutement de la snRNP U1 au site proximal. De plus, des interactions homotypiques et hétérotypiques entre molécules hnRNP H et hnRNP A1 peuvent se former. Ce mécanisme est aussi appuyé par une étude qui a montré que des sites de liaison pour les protéines hnRNP H ou hnRNP A1 positionnés aux extrémités d'un intron peuvent coopérer et activer l'épissage. L'hypothèse retenue pour expliquer cet effet implique un rapprochement des sites d'épissage distants par l'intermédiaire d'interactions entre protéines liées à l'ARN (Martinez-Contreras et al., 2006). Une autre évidence très importante qui vient soutenir notre modèle est la distribution des sites potentiels de liaison pour hnRNP H et hnRNP A1 dans le génome humain. En effet, des études statistiques ont révélé une fréquence élevée de la séquence GGG, qui constitue un site potentiel de liaison

pour ces molécules, près des sites d'épissage 5' et 3' dans les introns (figure 11) (Nussinov, 1987;Nussinov, 1988;Nussinov, 1989;Engelbrecht *et al.*, 1992;Solovyev *et al.*, 1994;McCullough et Berget, 1997;Majewski et Ott, 2002;Yeo *et al.*, 2004). Cette distribution de motifs, près des jonctions introns/exons, permettrait une meilleure efficacité dans l'exclusion de certains exons par l'intermédiaire des protéines hnRNP H et hnRNP A1. De plus, la présence de séquences GGG à l'extrémité des introns permettrait une juxtaposition beaucoup plus précise des sites d'épissage distants et aiderait grandement à masquer l'utilisation de pseudos sites situés entre ceux-ci.

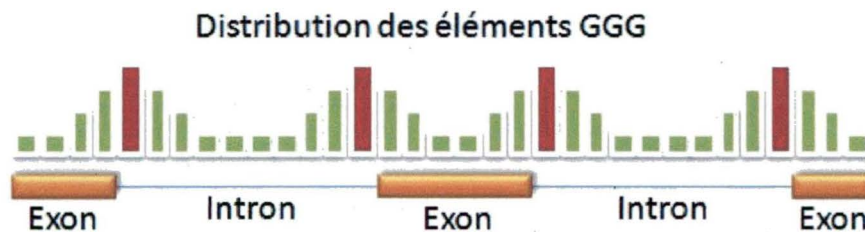


Figure 11. Distribution des sites potentiels pour hnRNP H et hnRNP A1 dans le génome humain. La localisation des éléments GGG est illustrée dans le schéma ci-dessus et indique une plus grande fréquence près des sites d'épissage (rectangles rouges) comparativement aux autres régions (rectangles verts). Données tirées de (Yeo et al., 2004).

Puisque nous retrouvons une quantité élevée de séquences GGG, dans le génome humain, il n'est pas envisageable que tous les exons qui sont bordés par cette disposition de nucléotides soient automatiquement exclus. En prenant en considération que les décisions d'épissage alternatif relèvent de la balance de plusieurs séquences régulatrices et facteurs, des niveaux très complexes de contrôle devraient moduler l'activité des sites de liaison pour hnRNP H et hnRNP A1. L'épissage alternatif du gène de la *thrombopoïétine* montre bien que les sites de liaison pour hnRNP H ne sont pas

nécessairement tous impliqués dans le choix d'un site d'épissage. En effet, l'exon 3 de ce gène contient deux sites d'épissage 3' en compétition qui comprennent dix motifs de liaison potentiels pour hnRNP H situés à proximité : seulement trois éléments riches en guanosines sont importants dans l'inhibition du site d'épissage 3' distal (Marcucci et al., 2007). Plusieurs hypothèses pourraient expliquer cette observation et nous aider à mieux comprendre comment l'activité des séquences liées par hnRNP H peut être contrôlée. Dans un premier temps, des sites potentiels de liaison pour hnRNP H pourraient ne pas être accessibles. D'une part, des facteurs situés à proximité de ces motifs pourraient être en compétition pour la liaison à la même section de l'unité d'épissage. D'autre part, il a aussi été reporté que les structures secondaires présentes dans l'ARN pré-messager peuvent diminuer l'accessibilité des facteurs d'épissage à leurs sites de liaison (Hiller et al., 2007). L'activité d'une protéine peut aussi être guidée par son état de phosphorylation. Par exemple, il a été documenté que la phosphorylation du facteur d'épissage Sam68 peut influencer sa capacité à moduler l'épissage alternatif de l'exon v5 du gène *CD44* (Matter et al., 2002). Chez le facteur hnRNP H, cette même modification post-traductionnelle peut s'effectuer par la protéine tyrosine kinase dans des granulocytes en apoptose ou en différenciation (Navakauskiene et al., 2004). Bien qu'aucune évidence n'ait établi les conséquences d'une telle phosphorylation, il est envisageable que l'activité d'épissage de cette protéine puisse être affectée.

Perspectives d'implications des autres régulateurs d'épissage

Étant donné que le facteur hnRNP F peut aussi lier des éléments riches en guanosines et qu'il contient les mêmes domaines protéiques que hnRNP H, nous avons postulé initialement que cette protéine pourrait aussi rapprocher des sites d'épissage

distants par l'intermédiaire d'interactions protéines : protéines. Cependant, aucune des séquences (HBS1, HBS2 et HBS3) n'a la capacité de former des complexes avec la protéine hnRNP F dans des expériences de gel de retardement (figure 4B, Chapitre 2 et figure non incluse), suggérant que nos motifs utilisés ne permettent pas la liaison à cette molécule. D'un autre côté, comme les protéines hnRNP H et hnRNP F peuvent former des hétérodimères (Chou et al., 1999), nous avons envisagé que hnRNP F pourrait tout de même participer à l'activité d'épissage par l'intermédiaire d'un recrutement via hnRNP H. Lorsque nous avons incubé notre unité d'épissage, contenant une combinaison de sites HBS1, dans un extrait nucléaire dont la concentration en protéines hnRNP H et hnRNP F a été diminuée, le pourcentage de produits proximaux augmentait (figure 4D, Chapitre 2). Cependant, l'ajout de protéines recombinantes hnRNP F à cet extrait nucléaire ne permettait pas de déplacer la sélection du site d'épissage 5' vers le distal, alors que la protéine recombinante hnRNP H augmentait son utilisation. Cette expérience nous suggère donc que la molécule hnRNP F n'est pas impliquée dans l'activité observée. Néanmoins, nous n'excluons pas la possibilité que ce facteur très similaire à hnRNP H puisse reproduire la stimulation du site d'épissage 5' distal dans des conditions où l'unité d'épissage que nous avons utilisée contiendrait des sites de liaison pour hnRNP F. Une étude structurale des portions correspondant aux qRRM de hnRNP F ont révélé que deux segments d'ARN peuvent lier une molécule constituée du qRRM 1 et qRRM 2 (Dominguez et Allain, 2006). Cette observation est en accord avec le modèle avec hnRNP A1 qui propose que le RRM 1 et RRM 2 lieraient des portions distinctes d'ARN afin de stabiliser les partenaires d'interaction. De plus, étant donné que des molécules hnRNP H et hnRNP F peuvent interagir, nous pourrions concevoir que ce complexe hétérotypique pourrait se former suite à la liaison de ces protéines et modifierait la sélection des sites

d'épissage 5'. Cette hypothèse d'une participation de hnRNP F pourrait aussi être étendue à d'autres protéines hnRNP qui possèdent des domaines protéiques similaires à hnRNP A1 et hnRNP H. Les régulateurs hnRNP G, P2, Q et R seraient d'excellents candidats puisqu'en plus de contenir des motifs de liaison à l'ARN et des GRD, ils ont tous été impliqués dans l'épissage (Martinez-Contreras et al., 2007). L'utilisation d'une telle variété de protéines hnRNP aurait l'avantage de créer un réseau très diversifié d'interactions protéines : protéines et donnerait plus de flexibilité à une cellule dans la modulation de l'épissage alternatif. Une telle complexité serait vraisemblablement accompagnée de mécanismes de contrôle rigoureux. La participation d'une troisième molécule, s'intercalant entre des partenaires d'interaction, pourrait moduler la formation des complexes. À titre d'exemple, une étude qui a révélé un réseau d'interactions entre hnRNP H et hnRNP A1 (figure 12) nous présente plusieurs partenaires communs pour les facteurs hnRNP H et hnRNP A1 (Jeronimo et al., 2007). Parmi ceux-ci, nous retrouvons hnRNP K, hnRNP M, hnRNP A2 et KHSRP qui sont tous des facteurs ayant été impliqués dans l'épissage et qui pourraient être d'excellents candidats.

Application de notre modèle

Le mécanisme que nous avons proposé pourrait permettre de mieux comprendre des événements d'épissage alternatif. Dans certaines situations, notre modèle peut être appliqué afin d'expliquer l'activité de régulateurs d'épissage. Par exemple, les facteurs Nova-1 et Nova-2 ont une distribution de sites de liaison qui est apparentée à celle des protéines hnRNP A1 et hnRNP H. En effet, une fréquence élevée de ces motifs est retrouvée près des sites d'épissage et il a été proposé que, dans certaines situations, la

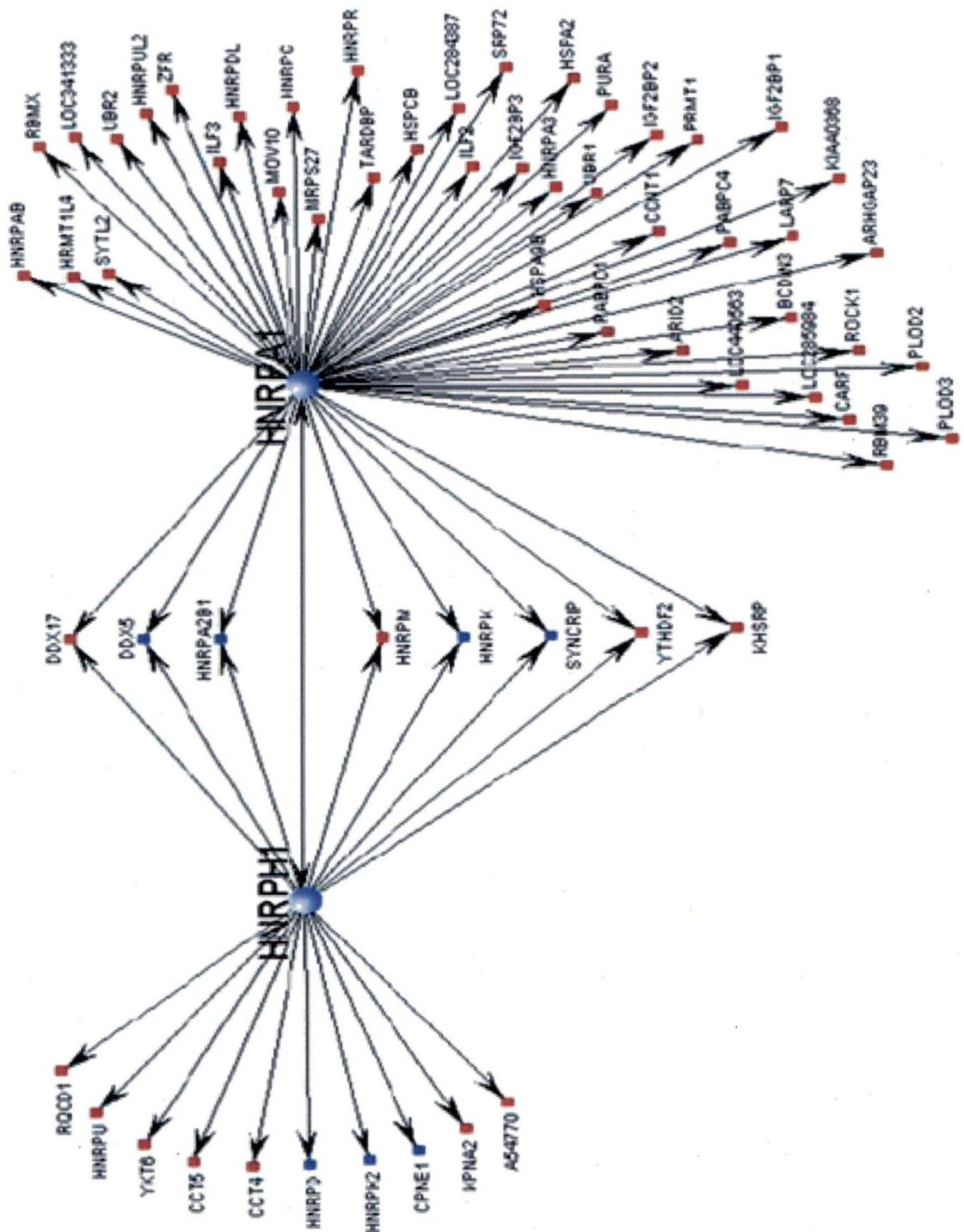


Figure 12. Réseau d'interactions des protéines hnRNP H et hnRNP A1. Suite à l'expression et l'immunoprécipitation des protéines hnRNP H et hnRNP A1, des analyses de spectrométrie de masse ont révélé plusieurs partenaires d'interaction pour ces deux facteurs d'épissage. Tiré de (Jeronimo et al., 2007). Les points bleus représentent des facteurs de maturation de l'ARN et les points rouges indiquent des facteurs impliqués dans le contrôle de la transcription.

formation de multimères rapprocherait un site d'épissage 5' à un site de branchement distant (Ule et al., 2006). Dans un autre contexte, la protéine Nova-1 régulerait l'inclusion d'un exon alternatif de son propre ARN pré-messager par l'intermédiaire d'un premier élément intronique et d'un second situé dans l'exon alternatif. Ces deux séquences contiennent des motifs de liaison pour la protéine Nova-1 et il a été proposé que la formation de multimères entre ces molécules masquerait l'utilisation des sites d'épissage et provoquerait l'exclusion de cet exon (Dredge et al., 2005).

CONCLUSIONS

Dans le cadre de notre étude portant sur la protéine hnRNP H, nous pouvons confirmer que ce facteur module l'épissage alternatif de façon similaire à hnRNP A1. En effet, nous avons montré que des combinaisons de motifs de liaison pour les facteurs hnRNP H et hnRNP A1 peuvent coopérer pour moduler la sélection des sites d'épissage 5'. De plus, nous avons observé que cette activité ne passe pas par une activation ou inactivation directe de ces sites d'épissage. Finalement, nous avons documenté que des interactions homotypiques et hétérotypiques entre molécules hnRNP H et hnRNP A1 peuvent se former dans les cellules vivantes. Ces évidences supportent fortement que hnRNP H peut stimuler l'épissage en rapprochant des sites d'épissage distants grâce à la formation d'interactions entre molécules liées à l'ARN. Afin de permettre une meilleure compréhension de ce modèle d'épissage, il serait très pertinent de s'attarder à des contextes naturels d'épissage alternatif gérés par ce type de mécanisme. Considérant que les protéines hnRNP H et hnRNP A1 peuvent moduler des décisions d'épissage de plusieurs ARN pré-messagers et que les façons d'y parvenir sont très diversifiées, il serait très intéressant de vérifier dans laquelle de ces situations notre modèle proposé est impliqué. L'importance d'une meilleure compréhension de ce mécanisme est accentuée par la nécessité de mieux prédire comment l'épissage des ARN pré-messagers s'effectue. Dans la mesure où des dérèglements dans des ratios d'isoformes peuvent engendrer plusieurs maladies graves, l'établissement de stratégies thérapeutiques doit passer par une meilleure vision de l'épissage alternatif des ARN pré-messagers.

REMERCIEMENTS

Je tiens en premier lieu à remercier mon directeur de recherche, le Dr Benoit Chabot, pour m'avoir transmis sa passion pour la recherche et enseigné que la persévérance mène inmanquablement vers la réussite.

Merci à mes nombreux collègues de travail, dont Timothée Revil, David Faucher et Martin Toussaint pour les captivantes discussions du midi et leurs appuis dans mes études doctorales.

Un merci tout spécial à Johanne Toutant pour son aide inestimable dans l'accomplissement de mes travaux et pour sa présence de tous les jours.

Finalement, je tiens à remercier ma femme Vicky Grondin pour ses nombreux encouragements et pour son appui dans la réalisation de mes projets futurs.

RÉFÉRENCES

- Alkan, S. A., Martincic, K., et Milcarek, C. 2006. The hnRNPs F and H2 bind to similar sequences to influence gene expression. *Biochem. J.* **393**: 361-371.
- Arhin, G. K., Boots, M., Bagga, P. S., Milcarek, C., et Wilusz, J. 2002. Downstream sequence elements with different affinities for the hnRNP H/H' protein influence the processing efficiency of mammalian polyadenylation signals. *Nucleic Acids Res.* **30**: 1842-1850.
- Berglund, J. A., Abovich, N., et Rosbash, M. 1998. A cooperative interaction between U2AF65 and mBBP/SF1 facilitates branchpoint region recognition. *Genes Dev.* **12**: 858-867.
- Black, D. L. 2003. Mechanisms of alternative pre-messenger RNA splicing. *Annu. Rev. Biochem.* **72**: 291-336.
- Blanchette, M. et Chabot, B. 1997. A highly stable duplex structure sequesters the 5' splice site region of hnRNP A1 alternative exon 7B. *RNA* **3**: 405-419.
- Blanchette, M. et Chabot, B. 1999. Modulation of exon skipping by high-affinity hnRNP A1-binding sites and by intron elements that repress splice site utilization. *EMBO J.* **18**: 1939-1952.
- Boise, L. H., Gonzalez-Garcia, M., Postema, C. E., Ding, L., Lindsten, T., Turka, L. A., Mao, X., Nunez, G., et Thompson, C. B. 1993. bcl-x, a bcl-2-related gene that functions as a dominant regulator of apoptotic cell death. *Cell* **74**: 597-608.
- Boutz, P. L., Chawla, G., Stoilov, P., et Black, D. L. 2007. MicroRNAs regulate the expression of the alternative splicing factor nPTB during muscle development. *Genes Dev.* **21**: 71-84.

- Brow, D. A. 2002. Allosteric cascade of spliceosome activation. *Annu. Rev. Genet.* **36**: 333-360.
- Buratti, E., Baralle, M., De Conti, L., Baralle, D., Romano, M., Ayala, Y. M., et Baralle, F. E. 2004. hnRNP H binding at the 5' splice site correlates with the pathological effect of two intronic mutations in the NF-1 and TSHbeta genes. *Nucleic Acids Res.* **32**: 4224-4236.
- Buvoli, M., Cobianchi, F., Bestagno, M. G., Mangiarotti, A., Bassi, M. T., Biamonti, G., et Riva, S. 1990a. Alternative splicing in the human gene for the core protein A1 generates another hnRNP protein. *EMBO J.* **9**: 1229-1235.
- Buvoli, M., Cobianchi, F., Biamonti, G., et Riva, S. 1990b. Recombinant hnRNP protein A1 and its N-terminal domain show preferential affinity for oligodeoxynucleotides homologous to intron/exon acceptor sites. *Nucleic Acids Res.* **18**: 6595-6600.
- Caputi, M. et Zahler, A. M. 2001. Determination of the RNA binding specificity of the heterogeneous nuclear ribonucleoprotein (hnRNP) H/H'/F/2H9 family. *J. Biol. Chem.* **276**: 43850-43859.
- Caputi, M. et Zahler, A. M. 2002. SR proteins and hnRNP H regulate the splicing of the HIV-1 tev-specific exon 6D. *EMBO J.* **21**: 845-855.
- Cartegni, L., Maconi, M., Morandi, E., Cobianchi, F., Riva, S., et Biamonti, G. 1996. hnRNP A1 selectively interacts through its Gly-rich domain with different RNA-binding proteins. *J. Mol. Biol.* **259**: 337-348.
- Chabot, B., Blanchette, M., Lapierre, I., et La Branche, H. 1997. An intron element modulating 5' splice site selection in the hnRNP A1 pre-mRNA interacts with hnRNP A1. *Mol. Cell. Biol.* **17**: 1776-1786.

- Chabot, B., LeBel, C., Hutchison, S., Nasim, F. H., et Simard, M. J. 2003. Heterogeneous nuclear ribonucleoprotein particle A/B proteins and the control of alternative splicing of the mammalian heterogeneous nuclear ribonucleoprotein particle A1 pre-mRNA. *Prog. Mol. Subcell. Biol.* **31**: 59-88.
- Chou, M. Y., Rooke, N., Turck, C. W., et Black, D. L. 1999. hnRNP H is a component of a splicing enhancer complex that activates a c-src alternative exon in neuronal cells. *Mol. Cell. Biol.* **19**: 69-77.
- Cloutier, P., Toutant, J., Shkreta, L., Goekjian, S., Revil, T., et Chabot, B. 2008. Antagonistic effects of the SRp30c protein and cryptic 5' splice sites on the alternative splicing of the apoptotic regulator Bcl-x. *J. Biol. Chem.*
- Coles, J. L., Hallegger, M., et Smith, C. W. 2009. A nonsense exon in the Tpm1 gene is silenced by hnRNP H and F. *RNA* **15**: 33-43.
- Crawford, J. B. et Patton, J. G. 2006. Activation of alpha-tropomyosin exon 2 is regulated by the SR protein 9G8 and heterogeneous nuclear ribonucleoproteins H and F. *Mol. Cell. Biol.* **26**: 8791-8802.
- Didiot, M. C., Tian, Z., Schaeffer, C., Subramanian, M., Mandel, J. L., et Moine, H. 2008. The G-quartet containing FMRP binding site in FMR1 mRNA is a potent exonic splicing enhancer. *Nucleic Acids Res.* **36**: 4902-4912.
- Ding, J., Hayashi, M. K., Zhang, Y., Manche, L., Krainer, A. R., et Xu, R. M. 1999. Crystal structure of the two-RRM domain of hnRNP A1 (UP1) complexed with single-stranded telomeric DNA. *Genes Dev.* **13**: 1102-1115.
- Dominguez, C. et Allain, F. H. 2006. NMR structure of the three quasi RNA recognition motifs (qRRMs) of human hnRNP F and interaction studies with Bcl-x G-tract RNA: a novel mode of RNA recognition. *Nucleic Acids Res.* **34**: 3634-3645.

- Dredge, B. K., Stefani, G., Engelhard, C. C., et Darnell, R. B. 2005. Nova autoregulation reveals dual functions in neuronal splicing. *EMBO J.* **24**: 1608-1620.
- Engelbrecht, J., Knudsen, S., et Brunak, S. 1992. G+C-rich tract in 5' end of human introns. *J. Mol. Biol.* **227**: 108-113.
- Expert-Bezancon, A., Sureau, A., Durosay, P., Salesse, R., Groeneveld, H., Lecaer, J. P., et Marie, J. 2004. hnRNP A1 and the SR proteins ASF/SF2 and SC35 have antagonistic functions in splicing of beta-tropomyosin exon 6B. *J. Biol. Chem.* **279**: 38249-38259.
- Fairbrother, W. G., Yeh, R. F., Sharp, P. A., et Burge, C. B. 2002. Predictive identification of exonic splicing enhancers in human genes. *Science* **297**: 1007-1013.
- Faustino, N. A. et Cooper, T. A. 2003. Pre-mRNA splicing and human disease. *Genes Dev.* **17**: 419-437.
- Fisette, J. F., Michelle, L., Revil, T., et Chabot, B. 2009. [Guiding and integrating to control and diversify splicing]. *Med Sci (Paris)* **25**: 175-180.
- Garneau, D., Revil, T., Fisette, J. F., et Chabot, B. 2005. Heterogeneous nuclear ribonucleoprotein F/H proteins modulate the alternative splicing of the apoptotic mediator Bcl-x. *J. Biol. Chem.* **280**: 22641-22650.
- Goren, A., Ram, O., Amit, M., Keren, H., Lev-Maor, G., Vig, I., Pupko, T., et Ast, G. 2006. Comparative analysis identifies exonic splicing regulatory sequences--The complex definition of enhancers and silencers. *Mol. Cell* **22**: 769-781.
- Graveley, B. R. 2000. Sorting out the complexity of SR protein functions. *RNA* **6**: 1197-1211.
- Graveley, B. R. 2001. Alternative splicing: increasing diversity in the proteomic world. *Trends Genet.* **17**: 100-107.

- Graveley, B. R. 2005. Mutually exclusive splicing of the insect Dscam pre-mRNA directed by competing intronic RNA secondary structures. *Cell* **123**: 65-73.
- Graveley, B. R., Hertel, K. J., et Maniatis, T. 2001. The role of U2AF35 and U2AF65 in enhancer-dependent splicing. *RNA* **7**: 806-818.
- Han, K., Yeo, G., An, P., Burge, C. B., et Grabowski, P. J. 2005. A combinatorial code for splicing silencing: UAGG and GGGG motifs. *PLoS Biol.* **3**: e158.
- Hanamura, A., Caceres, J. F., Mayeda, A., Franza, B. R., Jr., et Krainer, A. R. 1998. Regulated tissue-specific expression of antagonistic pre-mRNA splicing factors. *RNA* **4**: 430-444.
- Hiller, M., Zhang, Z., Backofen, R., et Stamm, S. 2007. Pre-mRNA secondary structures influence exon recognition. *PLoS Genet.* **3**: e204.
- Honda, K., Yamada, T., Seike, M., Hayashida, Y., Idogawa, M., Kondo, T., Ino, Y., et Hirohashi, S. 2004. Alternative splice variant of actinin-4 in small cell lung cancer. *Oncogene* **23**: 5257-5262.
- Honore, B., Baandrup, U., et Vorum, H. 2004. Heterogeneous nuclear ribonucleoproteins F and H/H' show differential expression in normal and selected cancer tissues. *Exp. Cell Res.* **294**: 199-209.
- Honore, B., Rasmussen, H. H., Vorum, H., Dejgaard, K., Liu, X., Gromov, P., Madsen, P., Gesser, B., Tommerup, N., et Celis, J. E. 1995. Heterogeneous nuclear ribonucleoproteins H, H', and F are members of a ubiquitously expressed subfamily of related but distinct proteins encoded by genes mapping to different chromosomes. *J. Biol. Chem.* **270**: 28780-28789.
- Hutchison, S., LeBel, C., Blanchette, M., et Chabot, B. 2002. Distinct sets of adjacent heterogeneous nuclear ribonucleoprotein (hnRNP) A1/A2 binding sites control 5'

- splice site selection in the hnRNP A1 mRNA precursor. *J. Biol. Chem.* **277**: 29745-29752.
- Jacquenot, S., Mereau, A., Bilodeau, P. S., Damier, L., Stoltzfus, C. M., et Branlant, C. 2001. A second exon splicing silencer within human immunodeficiency virus type 1 tat exon 2 represses splicing of Tat mRNA and binds protein hnRNP H. *J. Biol. Chem.* **276**: 40464-40475.
- Jeronimo, C., Forget, D., Bouchard, A., Li, Q., Chua, G., Poitras, C., Therien, C., Bergeron, D., Bourassa, S., Greenblatt, J. *et al.* 2007. Systematic analysis of the protein interaction network for the human transcription machinery reveals the identity of the 7SK capping enzyme. *Mol. Cell* **27**: 262-274.
- Jurica, M. S. et Moore, M. J. 2003. Pre-mRNA splicing: awash in a sea of proteins. *Mol. Cell* **12**: 5-14.
- Kanopka, A., Muhlemann, O., et Akusjarvi, G. 1996. Inhibition by SR proteins of splicing of a regulated adenovirus pre-mRNA. *Nature* **381**: 535-538.
- Kashima, T., Rao, N., et Manley, J. L. 2007. An intronic element contributes to splicing repression in spinal muscular atrophy. *Proc. Natl. Acad. Sci.* **104**: 3426-3431.
- Keniry, M. A. 2000. Quadruplex structures in nucleic acids. *Biopolymers* **56**: 123-146.
- Kohtz, J. D., Jamison, S. F., Will, C. L., Zuo, P., Luhrmann, R., Garcia-Blanco, M. A., et Manley, J. L. 1994. Protein-protein interactions and 5'-splice-site recognition in mammalian mRNA precursors. *Nature* **368**: 119-124.
- Ladomery, M. R., Harper, S. J., et Bates, D. O. 2007. Alternative splicing in angiogenesis: the vascular endothelial growth factor paradigm. *Cancer Lett.* **249**: 133-142.

- Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., *et al.* 2001. Initial sequencing and analysis of the human genome. *Nature* **409**: 860-921.
- Lavigne, A., La Branche, H., Kornblihtt, A. R., et Chabot, B. 1993. A splicing enhancer in the human fibronectin alternate ED1 exon interacts with SR proteins and stimulates U2 snRNP binding. *Genes Dev.* **7**: 2405-2417.
- Liu, F. et Gong, C. X. 2008. Tau exon 10 alternative splicing and tauopathies. *Mol. Neurodegener.* **3**: 8.
- Majewski, J. et Ott, J. 2002. Distribution and characterization of regulatory elements in the human genome. *Genome Res.* **12**: 1827-1836.
- Makeyev, E. V., Zhang, J., Carrasco, M. A., et Maniatis, T. 2007. The MicroRNA miR-124 promotes neuronal differentiation by triggering brain-specific alternative pre-mRNA splicing. *Mol. Cell* **27**: 435-448.
- Marcucci, R., Baralle, F. E., et Romano, M. 2007. Complex splicing control of the human Thrombopoietin gene by intronic G runs. *Nucleic Acids Res.* **35**: 132-142.
- Martinez-Contreras, R., Cloutier, P., Shkreta, L., Fisette, J. F., Revil, T., et Chabot, B. 2007. hnRNP proteins and splicing control. *Adv. Exp. Med. Biol.* **623**: 123-147.
- Martinez-Contreras, R., Fisette, J. F., Nasim, F. U., Madden, R., Cordeau, M., et Chabot, B. 2006. Intronic binding sites for hnRNP A/B and hnRNP F/H proteins stimulate pre-mRNA splicing. *PLoS Biol.* **4**: e21.
- Massiello, A., Roesser, J. R., et Chalfant, C. E. 2006. SAP155 Binds to ceramide-responsive RNA cis-element 1 and regulates the alternative 5' splice site selection of Bcl-x pre-mRNA. *FASEB J.* **20**: 1680-1682.

- Matter, N., Herrlich, P., et König, H. 2002. Signal-dependent regulation of splicing via phosphorylation of Sam68. *Nature* **420**: 691-695.
- Matunis, E. L., Matunis, M. J., et Dreyfuss, G. 1993. Association of individual hnRNP proteins and snRNPs with nascent transcripts. *J. Cell Biol.* **121**: 219-228.
- Mauger, D. M., Lin, C., et Garcia-Blanco, M. A. 2008. hnRNP H and hnRNP F complex with Fox2 to silence fibroblast growth factor receptor 2 exon IIIc. *Mol. Cell. Biol.* **28**: 5403-5419.
- McCullough, A. J. et Berget, S. M. 1997. G triplets located throughout a class of small vertebrate introns enforce intron borders and regulate splice site selection. *Mol. Cell. Biol.* **17**: 4562-4571.
- Min, H., Chan, R. C., et Black, D. L. 1995. The generally expressed hnRNP F is involved in a neural-specific pre-mRNA splicing event. *Genes Dev.* **9**: 2659-2671.
- Mount, S. M., Pettersson, I., Hinterberger, M., Karmas, A., et Steitz, J. A. 1983. The U1 small nuclear RNA-protein complex selectively binds a 5' splice site in vitro. *Cell* **33**: 509-518.
- Nasim, F. U., Hutchison, S., Cordeau, M., et Chabot, B. 2002. High-affinity hnRNP A1 binding sites and duplex-forming inverted repeats have similar effects on 5' splice site selection in support of a common looping out and repression mechanism. *RNA* **8**: 1078-1089.
- Navakauskiene, R., Treigyte, G., Gineitis, A., et Magnusson, K. E. 2004. Identification of apoptotic tyrosine-phosphorylated proteins after etoposide or retinoic acid treatment. *Proteomics* **4**: 1029-1041.
- Ni, J. Z., Grate, L., Donohue, J. P., Preston, C., Nobida, N., O'Brien, G., Shiue, L., Clark, T. A., Blume, J. E., et Ares, M., Jr. 2007. Ultraconserved elements are associated with

- homeostatic control of splicing regulators by alternative splicing and nonsense-mediated decay. *Genes Dev.* **21**: 708-718.
- Nussinov, R. 1987. (A)GGG(A), (A)CCC(A) and other potential 3' splice signals in primate nuclear pre-mRNA sequences. *Biochim. Biophys. Acta* **910**: 261-270.
- Nussinov, R. 1988. Conserved quartets near 5' intron junctions in primate nuclear pre-mRNA. *J. Theor. Biol.* **133**: 73-84.
- Nussinov, R. 1989. Conserved signals around the 5' splice sites in eukaryotic nuclear precursor mRNAs: G-runs are frequent in the introns and C in the exons near both 5' and 3' splice sites. *J. Biomol. Struct. Dyn.* **6**: 985-1000.
- Pan, Q.,Shai, O.,Lee, L. J.,Frey, B. J., et Blencowe, B. J. 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat. Genet.* **40**: 1413-1415.
- Paradis, C.,Cloutier, P.,Shkreta, L.,Toutant, J.,Klarskov, K., et Chabot, B. 2007. hnRNP I/PTB can antagonize the splicing repressor activity of SRp30c. *RNA* **13**: 1287-1300.
- Paronetto, M. P.,Achsel, T.,Massiello, A.,Chalfant, C. E., et Sette, C. 2007. The RNA-binding protein Sam68 modulates the alternative splicing of Bcl-x. *J. Cell Biol.* **176**: 929-939.
- Pennisi, E. 2005. Why do humans have so few genes? *Science* **309**: 80.
- Pinol-Roma, S.,Choi, Y. D.,Matunis, M. J., et Dreyfuss, G. 1988. Immunopurification of heterogeneous nuclear ribonucleoprotein particles reveals an assortment of RNA-binding proteins. *Genes Dev.* **2**: 215-227.

- Pozzoli, U., Sironi, M., Cagliani, R., Comi, G. P., Bardoni, A., et Bresolin, N. 2002. Comparative analysis of the human dystrophin and utrophin gene structures. *Genetics* **160**: 793-798.
- Revil, T., Toutant, J., Shkreta, L., Garneau, D., Cloutier, P., et Chabot, B. 2007. Protein kinase C-dependent control of Bcl-x alternative splicing. *Mol. Cell. Biol.* **27**: 8431-8441.
- Robberson, B. L., Cote, G. J., et Berget, S. M. 1990. Exon definition may facilitate splice site selection in RNAs with multiple exons. *Mol. Cell. Biol.* **10**: 84-94.
- Romano, M., Marcucci, R., Buratti, E., Ayala, Y. M., Sebastio, G., et Baralle, F. E. 2002. Regulation of 3' splice site selection in the 844ins68 polymorphism of the cystathionine Beta -synthase gene. *J. Biol. Chem.* **277**: 43821-43829.
- Rosbash, M. et Seraphin, B. 1991. Who's on first? The U1 snRNP-5' splice site interaction and splicing. *Trends Biochem. Sci.* **16**: 187-190.
- Rosbach, O., Hung, L. H., Schreiner, S., Grishina, I., Heiner, M., Hui, J., et Bindereif, A. 2009. Auto- and crossregulation of the hnRNP L proteins by alternative splicing. *Mol. Cell. Biol.*
- Rowen, L., Young, J., Birditt, B., Kaur, A., Madan, A., Philipps, D. L., Qin, S., Minx, P., Wilson, R. K., Hood, L. *et al.* 2002. Analysis of the human neurexin genes: alternative splicing and the generation of protein diversity. *Genomics* **79**: 587-597.
- Simard, M. J. et Chabot, B. 2000. Control of hnRNP A1 alternative splicing: an intron element represses use of the common 3' splice site. *Mol. Cell. Biol.* **20**: 7353-7362.
- Simard, M. J. et Chabot, B. 2002. SRp30c is a repressor of 3' splice site utilization. *Mol. Cell. Biol.* **22**: 4001-4010.

- Sironi, M., Menozzi, G., Riva, L., Cagliani, R., Comi, G. P., Bresolin, N., Giorda, R., et Pozzoli, U. 2004. Silencer elements as possible inhibitors of pseudoexon splicing. *Nucleic Acids Res.* **32**: 1783-1791.
- Soares, L. M., Zanier, K., Mackereth, C., Sattler, M., et Valcarcel, J. 2006. Intron removal requires proofreading of U2AF/3' splice site recognition by DEK. *Science* **312**: 1961-1965.
- Solovyev, V. V., Salamov, A. A., et Lawrence, C. B. 1994. Predicting internal exons by oligonucleotide composition and discriminant analysis of spliceable open reading frames. *Nucleic Acids Res.* **22**: 5156-5163.
- Sorek, R. et Ast, G. 2003. Intronic sequences flanking alternatively spliced exons are conserved between human and mouse. *Genome Res.* **13**: 1631-1637.
- Sorek, R., Shemesh, R., Cohen, Y., Basechess, O., Ast, G., et Shamir, R. 2004. A non-EST-based method for exon-skipping prediction. *Genome Res.* **14**: 1617-1623.
- Sugnet, C. W., Kent, W. J., Ares, M., Jr., et Haussler, D. 2004. Transcriptome and genome conservation of alternative splicing events in humans and mice. *Pac. Symp. Biocomput.*: 66-77.
- Sun, H. et Chasin, L. A. 2000. Multiple splicing defects in an intronic false exon. *Mol. Cell. Biol.* **20**: 6414-6425.
- Swanson, M. S. et Dreyfuss, G. 1988. Classification and purification of proteins of heterogeneous nuclear ribonucleoprotein particles by RNA-binding specificities. *Mol. Cell. Biol.* **8**: 2237-2241.
- Ule, J., Stefani, G., Mele, A., Ruggiu, M., Wang, X., Taneri, B., Gaasterland, T., Blencowe, B. J., et Darnell, R. B. 2006. An RNA map predicting Nova-dependent splicing regulation. *Nature* **444**: 580-586.

- Venables, J. P.,Klinck, R.,Bramard, A.,Inkel, L.,Dufresne-Martin, G.,Koh, C.,Gervais-Bird, J.,Lapointe, E.,Froehlich, U.,Durand, M. *et al.* 2008a. Identification of alternative splicing markers for breast cancer. *Cancer Res.* **68**: 9525-9531.
- Venables, J. P.,Koh, C. S.,Froehlich, U.,Lapointe, E.,Couture, S.,Inkel, L.,Bramard, A.,Paquet, E. R.,Watier, V.,Durand, M. *et al.* 2008b. Multiple and specific mRNA processing targets for the major human hnRNP proteins. *Mol. Cell. Biol.* **28**: 6033-6043.
- Veraldi, K. L.,Arhin, G. K.,Martincic, K.,Chung-Ganster, L. H.,Wilusz, J., et Milcarek, C. 2001. hnRNP F influences binding of a 64-kilodalton subunit of cleavage stimulation factor to mRNA precursors in mouse B cells. *Mol. Cell. Biol.* **21**: 1228-1238.
- Wang, E. T.,Sandberg, R.,Luo, S.,Khrebtkova, I.,Zhang, L.,Mayr, C.,Kingsmore, S. F.,Schroth, G. P., et Burge, C. B. 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature* **456**: 470-476.
- Wang, Z.,Hoffmann, H. M., et Grabowski, P. J. 1995. Intrinsic U2AF binding is modulated by exon enhancer signals in parallel with changes in splicing activity. *RNA* **1**: 21-35.
- Wang, Z.,Rolish, M. E.,Yeo, G.,Tung, V.,Mawson, M., et Burge, C. B. 2004. Systematic identification and analysis of exonic splicing silencers. *Cell* **119**: 831-845.
- Wang, Z.,Xiao, X.,Van Nostrand, E., et Burge, C. B. 2006. General and specific functions of exonic splicing silencers in splicing control. *Mol. Cell* **23**: 61-70.
- Will, C. L.,Schneider, C.,MacMillan, A. M.,Katopodis, N. F.,Neubauer, G.,Wilm, M.,Luhrmann, R., et Query, C. C. 2001. A novel U2 and U11/U12 snRNP protein that associates with the pre-mRNA branch site. *EMBO J* **20**: 4536-4546.

- Wollerton, M. C., Gooding, C., Wagner, E. J., Garcia-Blanco, M. A., et Smith, C. W. 2004. Autoregulation of polypyrimidine tract binding protein by alternative splicing leading to nonsense-mediated decay. *Mol. Cell* **13**: 91-100.
- Wu, J. Y. et Maniatis, T. 1993. Specific interactions between proteins implicated in splice site selection and regulated alternative splicing. *Cell* **75**: 1061-1070.
- Yeo, G., Hoon, S., Venkatesh, B., et Burge, C. B. 2004. Variation in sequence and organization of splicing regulatory elements in vertebrate genes. *Proc. Natl. Acad. Sci.* **101**: 15700-15705.
- Yeo, G. W., Coufal, N. G., Liang, T. Y., Peng, G. E., Fu, X. D., et Gage, F. H. 2009. An RNA code for the FOX2 splicing regulator revealed by mapping RNA-protein interactions in stem cells. *Nat. Struct. Mol. Biol.*
- Zahler, A. M., Damgaard, C. K., Kjems, J., et Caputi, M. 2004. SC35 and heterogeneous nuclear ribonucleoprotein A/B proteins bind to a juxtaposed exonic splicing enhancer/exonic splicing silencer element to regulate HIV-1 tat exon 2 splicing. *J. Biol. Chem.* **279**: 10077-10084.
- Zhang, X. H. et Chasin, L. A. 2004. Computational definition of sequence motifs governing constitutive exon splicing. *Genes Dev.* **18**: 1241-1250.
- Zuo, P. et Maniatis, T. 1996. The splicing factor U2AF35 mediates critical protein-protein interactions in constitutive and enhancer-dependent splicing. *Genes Dev.* **10**: 1356-1368.